ELEMENTAL MASS QUANTIFICATION FROM SPECTRAL X-RAY RADIOGRAPHS AND FLUORESCENCE USING GAUSS-NEWTON AND DEEP LEARNING APPROACHES

A Dissertation
Presented to
The Academic Faculty

By

Wesley C. Gillis

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Woodruff School of Mechanical Engineering
Nuclear Engineering and Medical Physics Program

Georgia Institute of Technology

December 2020

ELEMENTAL MASS QUANTIFICATION FROM SPECTRAL X-RAY RADIOGRAPHS AND FLUORESCENCE USING GAUSS-NEWTON AND DEEP LEARNING APPROACHES

Thesis committee:

Dr. Anna Erickson
Nuclear and Radiological Engineering and Medical Physics
Georgia Institute of Technology

Dr. Derek Haas
Nuclear and Radiation Engineering
University of Texas at Austin

Dr. Steven Biegalski
Nuclear and Radiological Engineering and Medical Physics
Georgia Institute of Technology

Dr. Karl Pazdernik
Applied Statistics and Computational Modeling
Pacific Northwest National Laboratory

Dr. Nolan Hertel
Nuclear and Radiological Engineering and Medical Physics
Georgia Institute of Technology

Dr. Andrew Gilbert
Applied Radiation Detection
Pacific Northwest National Laboratory

Date approved: December 2, 2020

For my parents, Dan and Michelle.

ACKNOWLEDGMENTS

I would first like to thank my advisor Dr. Anna Erickson for allowing me to work in her lab and providing guidance and support. You have taught me so much about how to navigate the challenges of research and academia. Your strong leadership and drive are astounding and inspirational.

To my mentor at PNNL and member of my committee, Dr. Karl Pazdernik, thank you for all your support and investment in me. You were a constant and strong presence in my research, providing insight, knowledge, and drive. I am grateful for the time I spent learning from you as a statistician and as a researcher.

I would also like to thank Dr. Andrew Gilbert, a member of my committee and a mentor. You provided me with support and enthusiasm for the past two years straight. You held me to high expectations and made me into a better researcher. I enjoyed working with you and hope to work together again as colleagues.

Thank you to my advisory committee members Dr. Steve Biegalski, Dr. Nolan Hertel, and Dr. Derek Haas for their feedback, suggestions, and flexibility.

Another thank you to the PNNL members Dr. Ben McDonald, Dr. Rick Wittman, Dr. Deb Fagan, Dr. Alex Hagen, and Becky Jones for a warm welcome and an invaluable experience.

To my many colleagues at Georgia Tech, John Stooksbury, Dr. Joseph Harms, Dr. Luke Maloney, Dr. Andrew Conant, Arith Rajapakse, and Jake Inman, thank you for the scintillating discussion and help.

A big thank you to the Nuclear Nonproliferation International Safeguards Graduate Fellowship for making my research possible. This research was performed under appointment to the Nuclear Nonproliferation International Safeguards Fellowship Program sponsored by the National Nuclear Security Administrations Office of International Nuclear Safeguards (NA-241).

To my family who provided me undying support and love. Thank you to my parents Dan and Michelle. You have always encouraged me to work hard with strong integrity. You've been my biggest allies, and I am eternally grateful for your unconditional love. I do not know what I did to deserve you. Thank you to my brother Ryan for being my first role model and for cheering me on all these years. To Oma, Opa, Grandma, and Grandad, thank you for your love and for always supporting me; you are my loudest and favorite fans. To my Aunt Regan, you are my oldest and most beloved mentor in science and in learning; thank you for all you have taught me and done for me. And to my cousin Becca, thank you for your love and unmatched wit. I cherish all the memories and laughs we have shared together.

And finally, to my partner Dr. Amnon Ortoll-Bloch, I am so grateful for your support, guidance, and love. You have helped motivate and encourage me to be the best I can be. Thank you for your companionship and the inspiration you give me each day.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The goal of this thesis is to explore elemental mass quantification from spectral X-ray radiographs and X-ray fluorescence. This would provide a nondestructive technique to the IAEA for international safeguards. The entire work's setup is a 160 kVp X-ray beam incident on a powder and measured with a pixelated spectral CdTe photon detector. First, the work implements a partial-volume correction to an existing numerical approach. An alternative deep learning approach is presented using CNNs to regress elemental mass. The training dataset is generated with Monte Carlo and empirical detector characterization. An unsupervised deep learning approach is also explored for the simulation-to-experiment transformation. The method is tested on both simulation and experimental data. Lastly, X-ray fluorescence from the sample is measured with a second, out-of-beam spectral photon detector. Similarly, deep learning is used to regress elemental mass. This is done both from X-ray fluorescence alone and fused with the spectral radiographic data. The work provides new technology to the IAEA and shows how simulation can be used in deep learning where experimental data is scarce.

CHAPTER 1
INTRODUCTION

1.1 Spectral X-ray Radiography for International Safeguards

The landmark Treaty on the Non-Proliferation of Nuclear Weapons (NPT) began an international agreement intended to ensure the peaceful use of nuclear technology [1]. Among the 191 state parties that signed the NPT, five states are considered the nuclear-weapon states, namely the United States, Russia, the United Kingdom, France, and China. Under the NPT, these states share a goal toward both nuclear and total disarmament. The majority of the remaining non-nuclear-weapon state parties concluded a comprehensive safeguards agreement (CSA), allowing the International Atomic Energy Agency (IAEA) to apply safeguards to the state's nuclear material. These safeguards reflect two of the NPT's objectives: the prevention of the spread of nuclear weapons and nuclear weapon technology and the promotion of the peaceful use of nuclear energy.

The IAEA implements and maintains a range of safeguard measures to carry out its responsibilities enumerated in the NPT. Under a CSA, the IAEA implements safeguards in nuclear fuel cycle facilities, including uranium conversion and enrichment, fuel fabrication, power reactors, spent fuel stores, and reprocessing. Some of the implemented technologies perform destructive or non-destructive analysis on material samples from facilities [2]. Researchers have been tasked with developing new and innovative safeguards technologies to add to the IAEA toolkit to address emerging and future needs.

X-ray interrogation is a diverse category of analysis tools, a number of which are currently implemented as safeguards technology. Methods such as K-edge densitometry (KEDG) [3], hybrid K-edge densitometry (HKED) [4], and the combined procedure for uranium concentration and enrichment assay (CMPU) [5] use X-rays to measure uranium

and/or plutonium concentrations from solution. This is ideal for the uranyl and plutonium nitrate solutions produced during spent fuel reprocessing. To use this method for solid uranium samples, such as uranium oxide powder from a fuel fabrication facility, the solution samples must be prepared in a laboratory. Alternatively, the X-ray fluorescence analyzer (XRFA) semi-quantitatively determines relative elemental concentrations of various sample types [6]. While the XRFA does not have the same strict requirements on sample preparation, it has limited precision. A quantitative X-ray system that directly measures powder samples would bypass the additional dissolution step in some facilities.

Spectral X-ray radiography was recently presented as a new safeguards technology for elemental mass quantification [7]. Using pixelated, spectroscopic detectors, an emergent technology in X-ray detection, the system was designed to assay uranium oxide powder samples from a nuclear fuel fabrication facility. Other transmission-based X-ray systems only assay solutions because they require the sample to be homogeneous and uniform. The detectors measure the spectral transmission with high energy resolution, but they assume that the beam is uniformly attenuated by the sample across the face of the detector. This assumption is invalid for powder samples. Spectral X-ray radiography instead makes use of detectors that are both energy-sensitive and position-sensitive. The proposed method permits samples with spatial variation by additionally measuring the X-ray spectrum in each pixel individually.

Pixelated, spectroscopic detector technology has matured over the last 20 years, but spectral X-ray radiography is still relatively young and faces its own technical obstacles. One such challenge results from limited spatial resolution. The partial-volume effect occurs when the density of the target has a sharp spatial gradient across the area projected onto a single pixel. The mass calculation uses a physics model that assumes that each pixel measures a uniform material attenuation. Large errors can arise, for example, in pixels that only partially view the target. The phenomenon plagues other related techniques, such as X-ray computed tomography (CT) [8] and positron emission tomography [9]. Corrections

to the measurements are typically only designed to qualitatively improve image quality by suppressing artifacts. An effective partial-volume correction for spectral X-ray radiography would result in improved elemental mass accuracy in all pixels, including those that measure partial attenuation.

One of the most challenging aspects of this system is the quantifying the total mass of each element compared to relative concentrations. While some neutron-based safeguards technologies determine total fissile, uranium, or plutonium mass, there are no safeguards that use X-ray interrogation to determine total elemental mass. Beyond the context of safeguards, X-rays have been used to quantify effective atomic number [10], electron density [11], and relative elemental concentration [12]; however, it is a nontrivial task to subsequently determine absolute quantities which often requires sample standards for calibration, additional types of measurements, or a high confidence in system geometry. Absolute quantification is a large obstacle for X-ray interrogation techniques like spectral X-ray radiography.

The accurate quantification of total elemental mass is an ongoing design challenge for spectral X-ray radiography. To calculate mass, Gilbert *et al.* [7] took an intermediate step to calculate the elemental areal density, or the mass per unit area, viewed by each pixel. The mass was then calculated from areal density using the distance between the sample and detector face, the distance between the source and the detector, the pixel size, and the sample thickness. The accuracy of the calculated mass is, therefore, highly dependent on the confidence of these geometric parameters. A technique that accurately calculates mass from measurements under systematic and statistical uncertainty would be beneficial to the IAEA. Such a system would be more affordable by reducing the construction and maintenance costs associated with high-precision instrumentation. It would also facilitate usability by relaxing the requirements of the operator.

1.2 Research Focus and Contributions

Many of the challenges associated with spectral X-ray radiography involve the computation of masses from measurement. The primary focus of this thesis is the exploration of computational techniques that address these obstacles, including optimization and deep learning. Designing for the use in safeguards would bring improvements in forms like usability, reduction in cost, or better quantitative performance. With these improvements, the technology would become even more suitable for safeguards applications.

The application of deep learning to this problem also presents an alternative methodology for computations from pixelated, spectroscopic X-ray detectors, particularly hyperspectral X-ray imaging detectors. Concepts from computer vision are used to exploit the rich spatial and spectral information available in the data. This work also shows how deep learning facilitates multi-modal fusion of this complicated measurement with the data from a different type of sensor that was measured simultaneously, specifically the fusion of spectral X-ray radiography and energy-dispersive X-ray fluorescence (XRF).

The work relies on simulation to implement deep learning that is to be used on experimental measurement. An important contribution of this thesis is the coupling of experiment and simulation through deep learning. Simulation benefits from perfect system knowledge, can explore a wide range of system parameters, and is cheap to generate. Deep learning is used on unpaired simulation outputs and experimental measurements for transferring simulation knowledge to the experimental domain and to model the translation between simulation and experiment. This method applies to nuclear security and safeguards techniques like X-ray radiography and spectroscopy where experiments are more costly than simulation or system information is limited. The impact reaches fields such as nondestructive testing and medical imaging by using deep learning to couple simulation and measurement. Coupling simulation and experiment together has exciting opportunities for the application of deep learning to engineering and the sciences.

1.3 Objectives and Structure

This work aims to contribute to the design of a spectral X-ray radiography system that quantifies the elemental mass of a high-Z powder sample. The intended result would show improved accuracy and robustness to systematic uncertainty. The study was guided by the following objectives:

1. To develop a partial-volume correction to the analytical approach
2. To determine how well deep supervised learning predicts elemental mass from simulated spectral X-ray radiography data
3. To determine if the mass prediction improves with the additional of energy-dispersive XRF via multi-modal fusion
4. To explore deep unsupervised learning for simulation/experiment domain translation and transfer learning of mass quantification

This first chapter introduced the overall context of the thesis, technology intended for international nuclear safeguards. The specific technology of interest is spectral X-ray radiography using an X-ray tube source and a hyperspectral X-ray imaging detector. Spatial resolution and absolute mass quantification limit the state-of-the-art systems. The main contributions of the work are delineated, followed by the specific objectives that guided the work.

The second chapter outlines the theoretical foundation of the thesis. The important technical information is a combination of the established theory and the presentation and critical review of the state-of-the-art. The basics of X-ray interrogation methods are first discussed, including types of sources, X-ray interaction mechanisms, and detectors. This is followed by a more detailed description of transmission, fluorescence, and multi-modal methods for interrogation intended to quantify one or multiple features of a target. Next,

specific information about X-ray detection with pixelated, spectroscopic detectors is highlighted, such as detection with semiconductors, detector materials, polarization in CdTe, pixelation, and commercially available detectors. The chapter concludes the following pertinent aspects of deep learning: deep learning architectures, tasks, training, and evaluation.

Chapter 3 details the underlying methods and materials that are at the core of this work. First described are the physical models used in algorithms, radiation transport calculations, and detector response modeling. Second, the details of the Monte Carlo simulations that are common to all chapters are listed. Finally, the experimental designs and procedures are presented.

Chapters 4 marks the first of three chapters describing methods, results, discussion, and conclusions of the three major topics explored in this thesis. Chapter 4 addresses the first objective described by implementing a partial-volume correction to the algorithm presented in [7]. It is similarly applied to uranium dioxide powder using a Gauss-Newton, reduced Hessian optimization to fit the data to a nonlinear physics model. Chapter 5 uses deep convolutional neural networks to predict elemental mass from spectral X-ray radiography alone and fused with X-ray fluorescence data, addressing the second and third objectives. Bismuth(III) oxide, gadolinium(III) oxide, and tungsten metal powders are used in place of uranium as surrogates to facilitate future experiments because of the restriction on obtaining uranium samples and because of its natural radioactivity. The fourth objective is explored in Chapter 6, where simulated and experimental bismuth(III) oxide measurements are coupled yielding two functionalities: a deep learning simulation-to-experiment response function and deep learning model that predicts elemental mass from experimental data.

The final chapter highlights the important conclusions and lessons learned from this thesis. The conclusions concerning each objective are first summarized. Then, the recommendations for continuing this work are discussed. Finally, this chapter enumerates the contributions of this thesis to the specific technique, quantitative imaging, and nondestructive testing.

CHAPTER 2
X-RAY INTERROGATION FOR MATERIAL DISCRIMINATION

The previous chapter outlined the motivation for elemental mass quantification using X-rays in nuclear safeguards. This technique belongs to a larger family of methods that use X-rays to probe materials at the atomic level. This chapter starts with a discussion of how X-rays reveal information about a material. Next, we review the detection technology at the core of this thesis: pixelated, spectral semiconductor detectors. We conclude the chapter with a survey of relevant material discrimination methods from multiple disciplines and application spaces. The section has a strong focus on spectral X-ray radiography.

2.1 A Brief Summary of X-ray Physics and Signatures

This section provides only a description of the relevant X-ray physics and signatures used throughout this work. Attix [13] provides a comprehensive description of X-ray, γ-ray, and charged particle physics.

2.1.1 X-ray Generation

X-rays are a type of ionizing radiation in the electromagnetic waves. Two different mechanisms are responsible for producing X-rays. The first is the deceleration of a charged particle, e.g. bremsstrahlung and synchrotron radiation. The second is the transition of electrons in inner atomic orbitals, e.g. fluorescence. Standard X-ray tubes exhibit both of these phenomena, which is evident in the X-ray energy spectrum in Figure 2.1. When an energetic electron beam strikes a metal target such as tungsten, the electrons decelerate and produce X-rays with a continuous energy distribution. The impinging electrons also ionize the tungsten by ejecting an orbital electron from an inner shell, leaving a vacancy. When an outer-shell electron falls to fill this vacancy, it emits an X-ray with energy equal to the

difference in energy levels. These X-rays are known as characteristic X-rays because each element has unique energy levels.

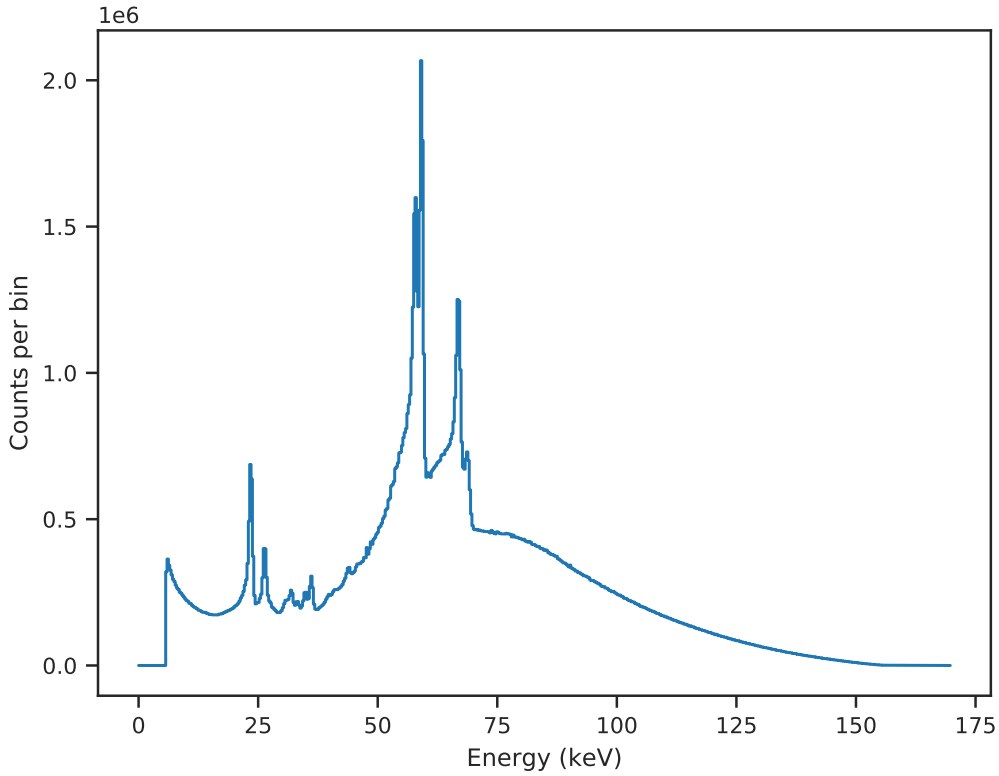

Figure 2.1: The energy spectrum shows a 160 kVp X-ray beam measured with a pixelated CdTe HEXITEC detector.

2.1.2 Types of Interaction Mechanisms

The utility of X-rays comes from how they interact with matter. We limit our discussion to the energies and, consequently, the interaction types typically used in X-ray interrogation. In photoelectric absorption, an incoming X-ray with energy $h\nu$ encounters an atomic electron with binding energy E_b. The X-ray disappears and the electron is ejected with kinetic energy $T = h\nu - E_b$.

Alternatively, the X-ray may scatter off an electron. Compton or incoherent scattering results a partial energy transfer. The reduced X-ray energy $h\nu'$ is determined by the angle

of scatter, and the electron exits with kinetic energy $T = h\nu - h\nu'$. Rayleigh or coherent scattering results in a small change in momentum but no energy transfer. Consequently, it's role is usually only significant at low X-ray energies.

2.1.3 Transmission Signatures

As an X-ray passes through a material, it has some probability of interacting with an atomic electron. The interaction probability and the likelihood of each interaction type are conditionally dependent on the X-ray's energy and the material's atomic number. The mass attenuation coefficient $\mu(E)$ is a value related to interaction probability through the material. Figure 2.2 shows total $\mu(E)$ (summed over all interaction types) as a function of X-ray energy for several elements. Generally, the absorption probability decreases as the X-ray energy increases. At low energies, the X-rays only have sufficient energy to eject the outer, more loosely bound electrons. The jump discontinuities in probability correspond to the minimum required energy to eject electrons from the next lower-energy shell, i.e. the X-ray's energy exceeds the electron's binding energy.

The energy-dependent features of the mass attenuation coefficients are elementally unique and observed via X-ray transmission. We first assume a narrow-beam geometry such that any X-ray that interacts in the material is removed from the beam. Before transmission, the X-rays are described by the energy-dependent fluence rate $\Phi_0(E)$. They impinge on a material composed of M materials where the mth material has areal density ρ_m and mass attenuation coefficient $\mu_m(E)$. The fluence rate after transmission $\Phi(E)$ is modeled by

$$\Phi(E) = \Phi_0(E) \exp\left[-(\boldsymbol{\mu}(E))^T \boldsymbol{\rho}\right]. \tag{2.1}$$

The term $\boldsymbol{\mu}(E) : \mathbb{R} \mapsto \mathbb{R}^M$ is the energy-dependent mass attenuation coefficient vector for the M materials, and $\boldsymbol{\rho} \in \mathbb{R}^M$ is the areal density vector for the M materials. Figure 2.3 compares measurements from a 160 kVp X-ray generator from a HEXITEC CdTe detector.

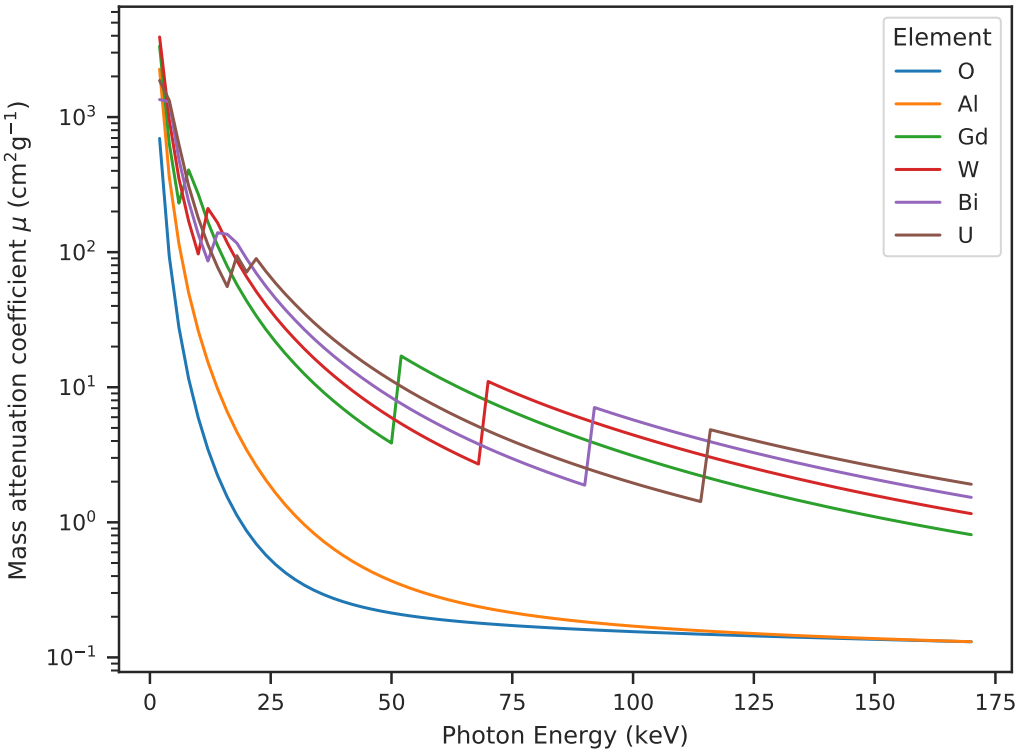

Figure 2.2: The XCOM mass attenuation coefficients of different elements the effect of atomic number and X-ray energy [14]. The discontinuities from the electron energy shells induce element-specific signatures in the X-ray energy distribution.

The raw beam measuring $\Phi_0(E)$ is plotted with the attenuated beams ($\Phi(E)$) from 1 mm W and Bi_2O_3 targets. The discontinuities in $\mu(E)$ that are visible in Figure 2.2 create prominent prominent spectral features in $\Phi(E)$.

X-ray radiography uses the X-ray transmission through a target to generate an image. Conventional radiography does not incorporate spectral signatures, using only spatially sensitive detectors. Dual-energy methods use the energy dependence either by combining two images, one taken with a low-energy X-ray beam and one taken with high-energy [15] or more recently by using photon counting detectors with two energy bins [10, 16]. K-edge subtraction imaging is a related technique where the X-ray energies are selected according to a particular element of interest [17]. Many methods using pixelated, energy-sensitive

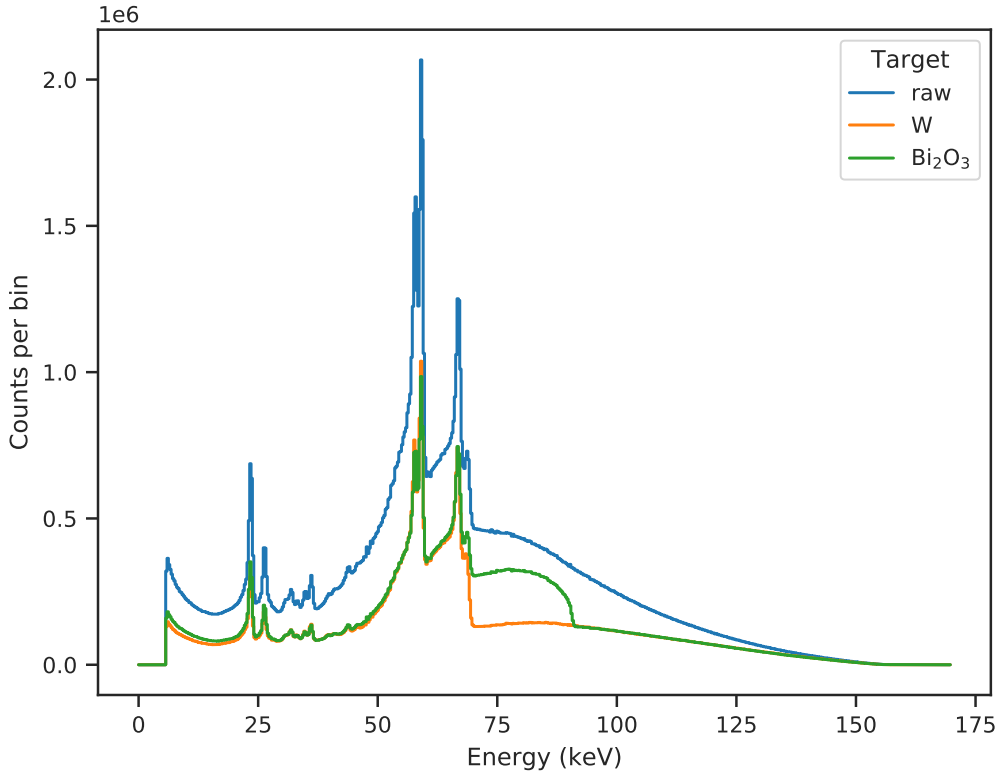

Figure 2.3: The transmission spectra through W and Bi_2O_3 are shown against the raw beam from a 160 kVp X-ray beam and 1 mm thick targets.

detectors have also been proposed, both for planar spectral radiography [18–21] and for spectral CT [22–25, 12, 26, 27]. CT is a technique that takes planar projections of the target at multiple views and subsequently attempts to reconstruct it in 3D.

2.1.4 Fluorescence Signatures

As X-rays interact with the target material, they eject atomic electrons from their orbitals. A vacancy left in an inner shell is filled when an outer shell electron drops down to this lower energy state. The electronic transition releases energy in the form of an X-ray equal to the difference in energy between the two states in a process called X-ray fluorescence (XRF). Each element has a unique electron orbital structure and consequently has a unique set of characteristic X-ray fluorescence energies. Figure 2.4 shows the signals measured

from various target materials interrogated by a 100 kVp X-ray beam. Compared to the transmission measurements in Figure 2.3, the XRF measurements are taken with the detector closer to the target and with the detector outside the beam. These signals are a small fraction of the intensity of the impinging beam.

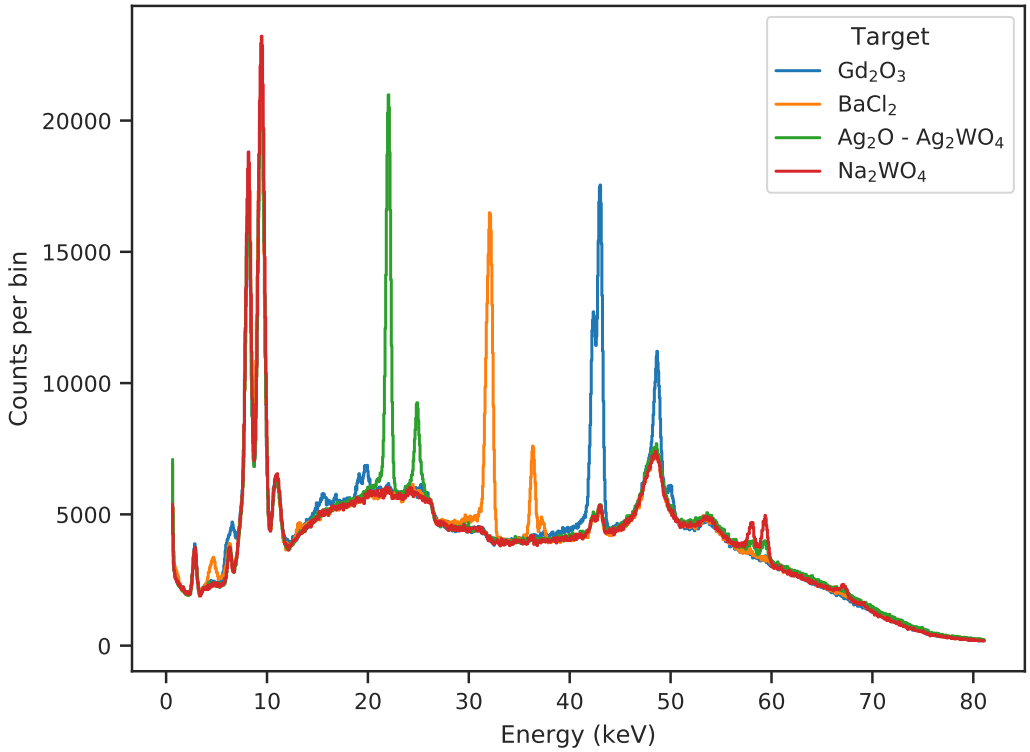

Figure 2.4: The XRF spectra of multiple powder targets were measured out-of-beam with an Amptek XR-100T CdTe.

For a target that is comparable to the mean-free path of an X-ray in that material, XRF signals suffer from self-attenuation. Qualitative measurements such as identifying elements is a standard task for an XRF system. For a quantitative task such as determining relative concentration, target geometry and homogeneity must be considered. Sample preparation may include dissolution or milling and pressing pellets.

Many XRF-based methods use synchrotrons as X-ray sources [28, 29] including those using pixelated, energy-sensitive detectors to create elemental mappings of a target's sur-

face [26, 30–32]. Several methods have been proposed, particularly for safeguards applications, for the quantification of metals and metal oxides [33–42]. These methods have limitations such as requiring a homogeneous target with well-known geometry or only providing surface information for large targets.

2.2 Pixelated, Spectroscopic X-ray Detectors

Some traditional X-ray detectors such as X-ray film, multi-wire proportional chambers, and flat panel detectors measure the spatial distribution of incoming X-rays. Such detectors are used in techniques such as X-ray radiography [43] and X-ray diffraction (XRD) [44]. Other detectors are instead designed to measure the energy distribution called spectral, spectroscopic, energy-dispersive, or energy-sensitive detectors. They are widely used in energy-dispersive X-ray fluorescence (EDXRF) [33], γ-ray spectroscopy [45], and neutron activation analysis [46]. This work uses a class of detectors that are capable of measuring both spatial and spectral distributions of incoming X-rays. Many names for these detectors are used in the literature such as energy-dispersive and energy-resolving imaging detectors, pixelated spectroscopic detectors, and many more variations. This section outlines many of the important features for the selection and operation of the detector used for material discrimination.

2.2.1 Semiconductor Diode Detectors

This information can be found in radiation detection references such as [47, 48] and quantum mechanics references such as [49].

Like all radiation detectors, the first step in measuring radiation with a semiconductor diode detector is the interaction of incoming radiation with the active detection volume. X-rays typically interact with electrons in the material, ejecting them, so they traverse the detector volume. This newly energetic electron ionizes the material along its path and slows down to a final stop. The amount of charge that accumulates in the volume from ionization

is roughly proportional to the total energy deposited by the electron; therefore, measuring energy deposition reduces to collecting the charge for each incoming event.

The underlying crystalline structure is responsible for the unique electrical properties of semiconductors. The lattice forms a periodic potential and results in nearly continuous, allowed energy bands separated by forbidden energy gaps. Electrons fill these energy bands according to the Fermi-Dirac distribution, parameterized by the Fermi level E_F. When at absolute zero, the highest-energy occupied states define the valence band, and the next vacant energy band is the conduction band.

Electrons can be excited into the conduction band by thermal noise or ionization of the material. In their place, they leave vacancies in the valence band called holes. The size of the forbidden gap E_g separating the these bands indicates the ease of excitation from the valence band to the conduction band. Semiconductors with small band gaps produce numerous electron-hole pairs in response to a radiation event; however, they are sensitive to thermal noise.

Semiconductors are often intentionally doped with impurities that impact the available electron states. An impurity in the lattice that has a leftover valence electron after covalently bonding to the neighboring atoms is a donor impurity present in n-type semiconductors. The extra electron is more loosely-bound, which creates an energy level in the forbidden gap near the conduction band. This proximity allows a higher concentration of electrons in the conduction band, making them the majority carrier. An impurity with one fewer valence electron than the surrounding atoms leaves a vacancy. The energy level of the vacancy sits just above the valence band because of the difference in orbital energies. Electrons fill these vacancies to satisfy the covalent bond and leave a hole in the valence band. These acceptor impurities, therefore, increase the concentration of holes, making them the majority carrier.

Semiconductor detectors are diodes, allowing current in only one direction. Many configurations of semiconductors achieve this feature. The p-n junction is a typical example where, for example, the end of a p-type semiconductor is heavily doped with donor im-

purities. A configuration that is relevant to this thesis is the Schottky diode, formed by a junction of a metal and a semiconductor.

A voltage applied in the reverse direction increases the resistivity, the potential difference, and the active detection volume. When a radiation event occurs in this volume, a burst of electron-hole pairs populates the detector. With a sufficient voltage bias, the electrons travel toward the anode in the conduction band, and the holes propagate toward the cathode in the valence band. The motion of the charge carriers induces a current until fully collected, which is integrated by the electronics.

2.2.2 Detector Material: Si, Ge, CdTe, CZT

This section highlights four semiconductor materials used as detectors: Si, Ge, CdTe, and CdZnTe (CZT) and provides the pros, cons, and typical uses of each.

Si

Si has a small bandgap which gives it excellent energy resolution. Due to its low atomic number, it has the lowest photon interaction probability. Consequently, Si is more transparent to γ-ray background and is typically limited to low-energy X-ray, alpha, and beta spectroscopy. One significant limitation on Si detectors is their small size due to fabrication restrictions. While some Si detectors do not require cooling by liquid nitrogen, others experience significant noise from leakage current without it.

Ge

Ge has an even smaller bandgap and higher atomic number than Si. They can also be grown with high purity to larger sizes; therefore, they are typically used for γ-ray spectroscopy. Their atomic number is much lower than other widely used γ-ray spectrometers such as NaI. High-purity Ge detectors require liquid nitrogen cooling during operation, and lithium drift Ge detectors must be cooled during operation and storage.

CdTe

CdTe has a much wider bandgap than either Si or Ge. This larger bandgap results in poorer energy resolution and lower sensitivity to thermal noise, though still higher than the best scintillators. CdTe can consequently be operated at room temperature. The high atomic numbers of Cd and Te provide high intrinsic efficiency but produce more significant escape peaks. CdTe can be grown to large sizes and is useful in imaging applications. Instability due to polarization is a well-known problem in CdTe that is caused by detrapping of deep acceptor sites located within the bandgap [50]. Additionally, the mobility of holes is significantly lower than the mobility of electrons in CdTe leading to visible hole-tailing of spectral peaks [51].

CZT

CdZnTe (CZT) has similar properties as CdTe as a detector. CZT, however, is more stable over time and does not suffer from polarization-induced degradation [52]. Higher resistivity compared to CdTe reduces the impact of leakage current, but the increased bandgap results in poorer energy resolution. The fabrication of CZT presents several difficulties that are prohibitive to growing large crystals [51].

2.2.3 Polarization and Bias Refresh in CdTe

Polarization is a major limitation of CdTe detectors. Consider a mostly p-type CdTe that has excess holes. It is attached to an anode such as In that has a high hole barrier and serves as an electron donor. Before a bias is applied, there is an overall negative space charge due to the electrons filling acceptor sites. After the bias is applied, the space charge changes over time. Electrons that are excited into the conduction band are collected at the anode. Because CdTe has issues with detrapping, electrons that fill acceptor sites near the cathode detrap and are subsequently removed at the anode. Trapping of electrons near the anode also occurs causing more buildup of negative space charge. The depletion region shrinks

away from the cathode and becomes localized around the anode [53].

The effect of polarization is typically worse for longer operation times, lower bias voltages, higher temperatures, and higher fluxes [54]. An alternative solution to reducing polarization in CdTe is called bias refreshed, proposed by Wilson *et al.* [52]. Seller *et al.* [55] demonstrated that polarization was successfully mitigated when measuring ^{241}Am for 6 hours by turning off the bias for 2 s every 60 s.

2.2.4 The Small Pixel Effect

The small pixel effect is used both for the pixelation of CdTe and CZT detectors, but also to improve detector performances. This effect is governed by the Shockley-Ramo Theorem for Induced Charged which is founded on two seminal works by Schockley [56] and Ramo [57]. It can be demonstrated by a detector with a single, large cathode on one side and pixelated anodes on the other. It also requires that the pixels are sufficiently smaller than the detector thickness. A consequence of the theorem is that the induced signal is highest from the motion of charged particles near the pixels. For pixelated anodes, nearly all the signal results from electron motion. For detectors like CdTe and CZT with poor hole mobility and carrier lifetimes, the small pixel effect improves spectral performance by only using the signal indued by electrons [58].

2.2.5 Charge Sharing

One challenge of this kind of pixelated detector is charge sharing. A radiation event that deposits energy in the detector generates a cloud of charge in the form of electrons and holes. The cloud primarily follows along field lines but is subject to diffusion. This diffusion can lead to the cloud spanning more than one pixel where each pixel measures only a fraction of the total energy. Veale *et al.* [59] found that charge sharing was less prominent with higher bias voltages and lower temperatures.

Two post-processing algorithms were presented by Veale *et al.* [54] to address charge

sharing using a HEXITEC CdTe detector. The charge sharing discrimination (CSD) algorithm removed events from pixels where a neighbor pixel also recorded an event. The charge sharing addition (CSA) algorithm instead summed the neighboring pixels with nonzero events and placed the result in the pixel with the highest measured signal. CSD had a lower efficiency but higher energy resolution. CSA retained more events at the expense of poorer energy resolution.

2.2.6 Hyperspectral and Multispectral X-ray Imaging

There is a recent distinction has been proposed based on how coarsely a detector measures energy [12]. Multispectral, single photon counting, or hybrid pixel detectors have crude energy resolution but can operate at higher count rates [60–63]. These function with a readout system that individually processes the signals from each pixel. Alternatively, hyperspectral detectors, like the one used in this work, have finer energy resolution [64]. They are limited by a maximum count rate lower than multispectral detectors. The readout system outputs frames containing the total energy deposited in each pixel within a small exposure time.

2.3 Material Discrimination Methods using Spectral X-ray Radiography

The various signals discussed above reveal information about the composition of a target. Here, we discuss the ways this information is extracted from these measurements. They differ in the nature of the target composition, the type of information being determined, the use of the source, the type of detector, the system geometry, *e.g.* planar vs. tomographic, and the algorithm used.

2.3.1 Basis Material Decomposition

The most widely researched method is basis material decomposition in response to the poor spectral behavior of imaging detectors. First introduced by Alvarez and Macovski [65], the approach uses dual-energy transmission measurements to solve for the composition of a

set of well-known basis materials. These basis materials are calibrated to the X-ray system and used to determine quantities such as volume fraction, effective atomic number, and effective density. These values, however, cannot be used to quantify mass. Brambilla *et al.* [21] demonstrated the successful extension of this method to more materials by using a mutlispectral detector and a Maximum Likelihood Estimation for decomposition. The mutlispectral detector permitted a higher-Z material than was previously possible.

Recent progress has also been made in basis material decomposition using multispectral detectors for CT. Both dual-energy [16] and spectral [25] CT methods have shown promising results, although there is skepticism about their quantification accuracy. First, assumptions about the forward projection model [19], beam hardening, and motion lead to artifacts and accuracy. Further, the accuracy depends on proper geometry calibration which is a challenging task still being explored [16].

One group of methods uses nonlinear least-squares to fit a physics model for material decomposition. Gilbert *et al.* [18] calculated areal density maps of U, O, and Gd then used geometric knowledge to calculate elemental mass. This work was presented in the context of international safeguards. Alternatively, Ducros *et al.* [19] used nonlinear least-squares for the material decomposition step in a material basis decomposition method using spectral CT. The context of this method was medical imaging.

2.3.2 Machine Learning Alternatives

The last few years have seen rising interest in machine learning techniques in place of traditional algorithmic methods. There has been much attention in material decomposition from post-reconstruction multispectral CT images. A major challenge in these and in many machine learning methods for physical sciences is determining the true label. This was noted by Lu *et al.* [66], where both simulated data and experimental data was tested. Their approach first used a mean-variation-median feature extractor then compared the performance of several supervised models. Simulated data has the benefit of having labels

but may not represent the experimental system. The experimental labels were determined using traditional material decomposition means. This begs the question: why use machine learning in the first place?

Jimenez [67] also presented a supervised approach, using a well-characterized geometry. They saw relatively poor performance, mostly limited by the size of the training set. This way of determining the ground truth for experiment was also explored by Chen and Li [68] who used two CNN architectures, the visual geometry group (VGG) network [69] and the deep residual network (ResNet) [70]. They trained both on a simulation dataset and on an experimental dataset from designed phantoms, splitting the images into smaller image patches. A phantom is an object used in medical imaging designed for testing and evaluation of an imaging system. They note that when using phantoms, it is difficult to ensure diversity in the training data and propose using realistic simulation training data.

One study done by Touch *et al.* [71] used a neural network for spectral distortion in spectral CT images, not for the material decomposition step. The study used simulation to create first ideal detector measurements then the spectrally distorted versions guided by experimental data. They trained the neural network to map the distorted versions to the idealized versions. Applying this to experimental data showed improvement in the subsequent material decomposition calculations. This work shows successful use of simulated training data on experimental data.

2.4 Multimodal Measurements with X-rays

In the previous section, several material decomposition methods were discussed. This section explores the use of multiple simultaneous measurements in the context of X-ray interrogation.

The use of multiple sensors in X-ray measurements has been of interest recently in many fields. Combinations include absorption-, phase-, and fluorescence-contrast mechanisms in hard X-ray imaging [72]; micro X-ray fluorescence (μ-XRF), micro X-ray diffrac-

tion (μ-XRD), and scanning transmission X-ray microscopy (STXM) computed tomography [73]; XRD and XRF [74]. In particular interest to our work which combines XRF and spectral X-ray radiography, Liotti *et al.* [31] combined XRF imaging and X-ray radiography, and Pereira *et al.* [75] combined a X-ray transmission microtomography with XRF microtomography (XRFCT).

Hybrid K-edge densitometry is a safeguards technique that uses spectral X-ray transmission and XRF [34]. At reprocessing plants, HKED determines uranium concentration (using transmission) and plutonium concentration (using XRF ratios and uranium concentration) [34]. It is also used for neptunium determination among fission products [2].

One machine learning approach for multimodal fusion was explored by Rahman *et al.* [76]. They fused digital images with XRF measurements of rock samples to quantify the elemental density. Feature extraction was performed on both input measurements individually. They compared the performance of linear regression and support-vector machine regression on density from the features.

CHAPTER 3

METHODS AND MATERIALS

3.1 Physical Models

3.1.1 Beer-Lambert Law for Spectral Transmission

The Beer-Lambert law of exponential attenuation describes the passage of X-rays through matter [13]. It is a governing principle of the physical model used for quantitative material discrimination from X-ray transmission. A radiant flux $\Phi_0(E)$ traveling through M materials is attenuated as

$$\Phi(E;\boldsymbol{\rho}) = \Phi_0(E)\exp\left[-(\boldsymbol{\mu}(E))^T\boldsymbol{\rho}\right] + \Phi_{\text{scatt}}(E) \tag{3.1}$$

where $\boldsymbol{\mu}(E) : \mathbb{R} \mapsto \mathbb{R}^M$ is the energy-dependent mass attenuation coefficient vector, and $\boldsymbol{\rho} \in \mathbb{R}^M$ is the areal density vector for M materials in the projection on the detector face. A component $\mu_m(E)$ has units cm^2/g, and a component ρ_m has units g/cm^2.The $\Phi_{\text{scatt}}(E)$ term describes the incoming flux of scattered X-rays. The radiant flux is spatially integrated over the face of the detector.

Assumption: The system has a narrow-beam geometry; therefore, the incoming scattering term $\Phi_{\text{scatt}}(E)$ is negligible, and the target described by $\boldsymbol{\rho}$ in view of the detector face is homogeneous.

This physical model has been used as the foundation for quantitative material discrimination methods. The goal is to solve for the density vector $\boldsymbol{\rho}$ which can then be used to find the mass vector $\boldsymbol{m}$.

3.1.2 Detector Response Functions

The next step in the physical model describes the detector's response to the incoming flux. The term $\Phi(E; \boldsymbol{\rho})$ in Equation 3.1 is the result flux due to transmission through $\boldsymbol{\rho}$. Incorporating a detector response function is applicable beyond transmission, including the X-ray fluorescence detector used in this work. We now consider a general, incoming X-ray flux $x(E)$.

We denote the energy E in Equation 3.1 of an X-ray impinging on the detector as *incoming energy*. We distinguish E from the *observed energy*, E', which is the energy of a pulse height measured in the detector due to energy deposition. The joint probability distribution of E and E', $r(E, E')$, is the detector response function. The observed energy distribution by the detector $y(E')$ in response to an incoming energy distribution $x(E)$ is

$$y(E') = \int_0^\infty r(E, E')x(E)dE. \tag{3.2}$$

In practice, detector response functions are discretized into energy bins. For a detector described by $r(E, E')$, the probability of the q^{th} incoming energy bin and the l^{th} observed energy bin is given by Equation 3.3, where E_q and E_l are the lower-energy bounds of bins q and l, respectively.

$$\int_{E_l}^{E_{l+1}} \int_{E_q}^{E_{q+1}} r(E, E')dEdE' = \int_{E_q}^{E_{q+1}} r_l(E)dE = r_{l,q} \tag{3.3}$$

The components $r_{l,q}$ form a matrix, $\boldsymbol{R} \in \mathbb{R}^{Q \times L}$, where Q and L are the number of incoming and observed energy bins, respectively. For a similarly discretized incoming flux $\boldsymbol{x} \in \mathbb{R}^{\mathrm{Q}}$ with components

$$x_q = \int_{E_q}^{E_{q+1}} x(E)dE, \tag{3.4}$$

the discretized detector response $\boldsymbol{y} \in \mathbb{R}^{\mathrm{L}}$ is approximated as a linear combination of the

columns of $\boldsymbol{R}$ by the entries of $\boldsymbol{x}$, as in Equation 3.5.

$$\boldsymbol{y} \approx \boldsymbol{R}\boldsymbol{x} \tag{3.5}$$

Assumption: The energy bins are sufficiently small such that

$$\int_{E_q}^{E_{q+1}} r_l(E)x(E)dE \approx \int_{E_q}^{E_{q+1}} r_l(E)dE \cdot \int_{E_q}^{E_{q+1}} x(E)dE \; \forall \, l, q.$$

HEXITEC CdTe Response Function

The spectral transmission detector models a HEXITEC with a CdTe volume. The response function was generated and experimentally validated by [77], shown in Figure 3.1. The each pixel of the detector measures a the incoming X-ray spectrum, and this response function represents the pixel-wise response to the incoming flux.

Amptek XR-100 CdTe Response Function

There was no available response function for an Amptek XR-100 CdTe detector. In addition to radiation transport theory, the response function generation uses a physics-based hole-tailing model described in [78]. In a detector of total thickness L, electron trapping length of λ_e, and hole trapping length of λ_h, charge collection efficiency as a function of interaction depth is

$$\eta(x) = \frac{\lambda_e}{L}\left(1 - e^{-\frac{x}{\lambda_e}}\right) + \frac{\lambda_h}{L}\left(1 - e^{-\frac{L-x}{\lambda_h}}\right) \tag{3.6}$$

where $L = 1$ mm, $\lambda_e = 13.2$ cm, and $\lambda_h = 0.8$ cm. The energy collected E_{collect} is from

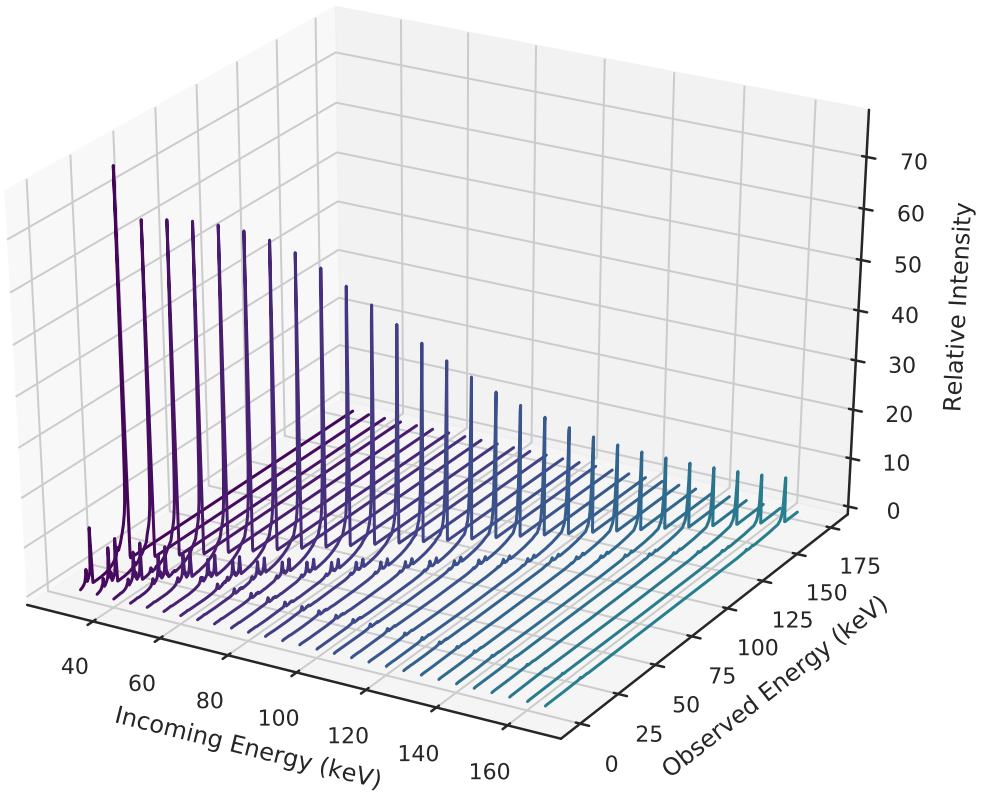

Figure 3.1: The pixel-wise detector response function of a HEXITEC CdTe was generated using MCNP [77].

the energy deposited E_{dep} as

$$E_{\text{collect}} = E_{\text{dep}} * \eta(x). \tag{3.7}$$

The radiation transport was performed on a simplified version of the Amptek XR-100 detector geometry, depicted in Figure 3.2. The detector volume consisted of a 5-mm $\times$ 5-mm $\times$ 1-mm CdTe parallelepiped with a density of 5.85 g/cm^3. A thin Be disk is centered 0.5 mm from the large face of the CdTe volume. The disk has a diameter of 7 mm and a thickness of 50 μm. Source photons were uniformly sampled on the face of a 7-mm disk located 1 mm from the center of the Be window. Each photon started with a momentum vector perpendicular to the Be window. The energies were uniformly sampled between 0 and 160 keV.

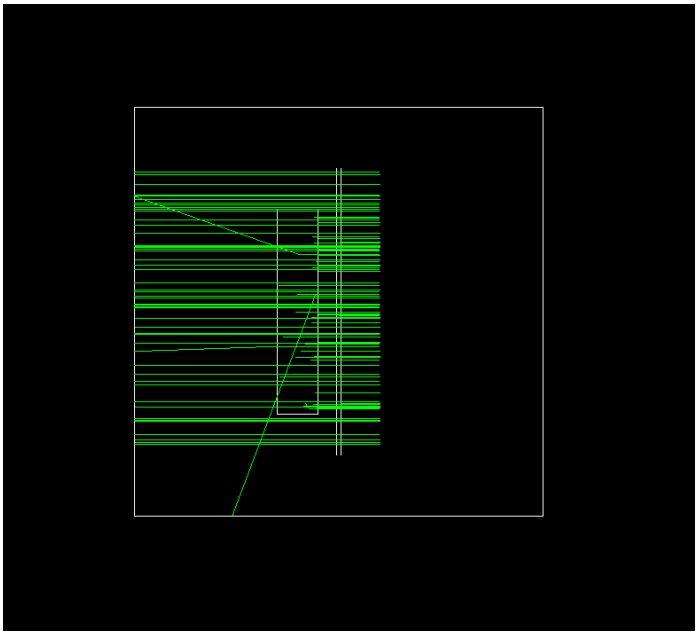

Figure 3.2: The geometry of the Amptek XR-100 CdTe was simplified for the simulations for the response function generation.

An event starts with a photon with energy E. If that photon interacts in the CdTe volume, it can eject an energetic electron. When the electron takes a step at depth x in the detector, it deposits energy E_{dep}, and the collected energy E_{collect} is calculated using Equation 3.7. The observed energy E' is the sum of the E_{collect} from each step of an event. At the end of each event, a 2D histogram is incremented with the pair (E, E').

The simulation yielded the energy deposition response to incoming photon energies. The response function was then spread by a Gaussian as described in [79]. The resulting response function is shown in Figure 3.3.

The final result is a physically motivated model of the observed energy distribution after transmission through a target. Explicitly, the L components of $\boldsymbol{y}$ are the responses in each observed energy bin l, given by

$$y_l = \sum_{q=0}^{Q-1} r_{l,q} \int_{E_q}^{E_{q+1}} x(E) dE. \tag{3.8}$$

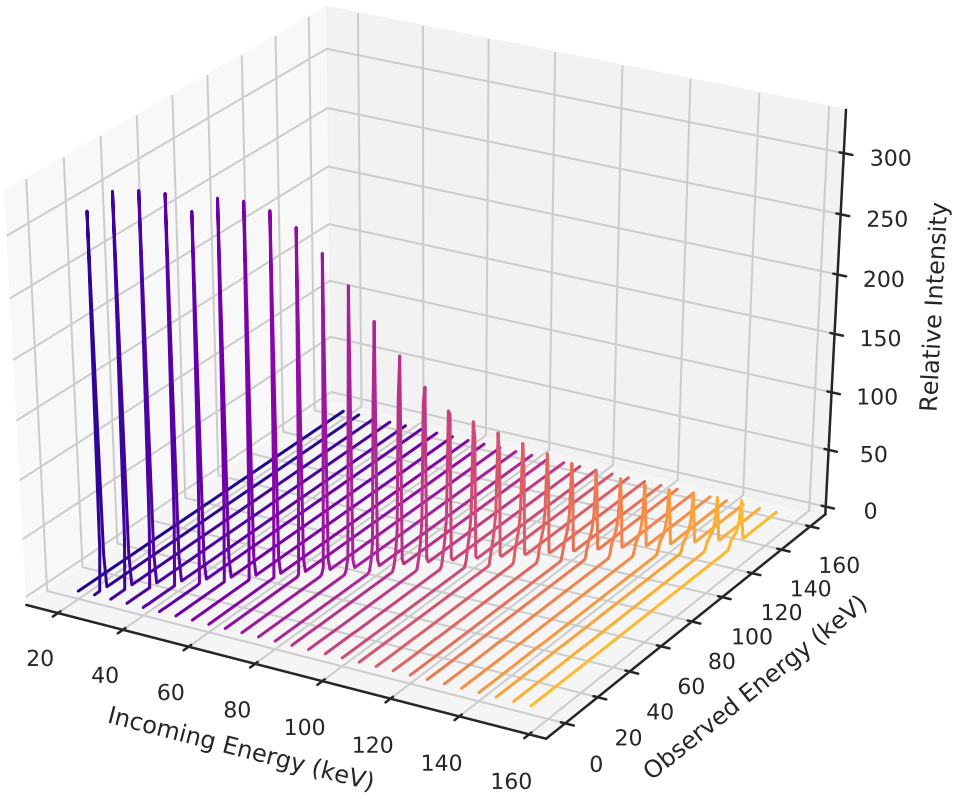

Figure 3.3: The detector response function of an Amptek XR-100 CdTe was generated using Geant4.

3.2 Monte Carlo Simulations

3.2.1 Tools and Resources

We used Geant4 10.4 to perform all the radiation transport in this work. Geant4 was developed by CERN and uses Monte Carlo to simulate physical processes using empirical and theoretical models [80, 81]. The computations used 15 64-core CPU nodes from the FoRCE Research Computing Environment cluster which was provided by the Georgia Institute of Technology.

The outputs of the Geant4 simulations were histograms in .root files. The ROOT package is a data analysis framework that was also developed my CERN. The Python bindings were used to read the histograms and perform data processing steps. Python was also used

to convert data to TFRecords to be used by a TensorFlow model.

3.2.2 Geometry Implementation and Randomization

The simulations used in this work share the similar geometries. The differences include powder target placement, powder composition, cuvette geometry, filter thickness and the presence of the XRF detector. Some simulations were generated to test and evaluate the partial-volume correction to the analytical mass quantification method. The remaining simulations were used for training, validation, and test data for the deep learning models. Figure 3.4a illustrates a standard example. The geometry can be subdivided into three distinct groupings: the source, the target, and the detectors. The parameters described here were used in Chapter 5 and Chapter 6; the parameters for Chapter 4 can be found within the chapter.

The source has both filtering and collimating components. The beam first encounters a 1.25-cm thick W collimator that has an aperture diameter of 1 cm. This is immediately followed by a 2- or 3-cm Al filter and a second 1.25 cm W collimator with a smaller aperture diameter of 2 mm. The source is distanced 205 cm from the transmission detector face. This design reduces the contribution from scattered X-rays at the detector, and the source X-rays are approximately parallel. Under these conditions, the assumptions on the physics model are more valid.

The distance of the target face to the detector face was 57 cm for Chapter 5 and Chapter 6 but differs in Chapter 4; however, they are all confined within the projection onto the transmission detector. The target is a acrylic cuvette containing a small, 1-mm thick volume filled with powder. The powder volume is voxelized into a grid of 100-μm $\times$ 100-μm $\times$ 1-mm voxels where the material in each voxel is randomly sampled from a distribution. This sampling scheme is described in Appendix C. This voxelization is finer than the spatial resolution of the pixelated detector.

In the simulations for deep learning, the cuvette's center and rotation in the global ge-

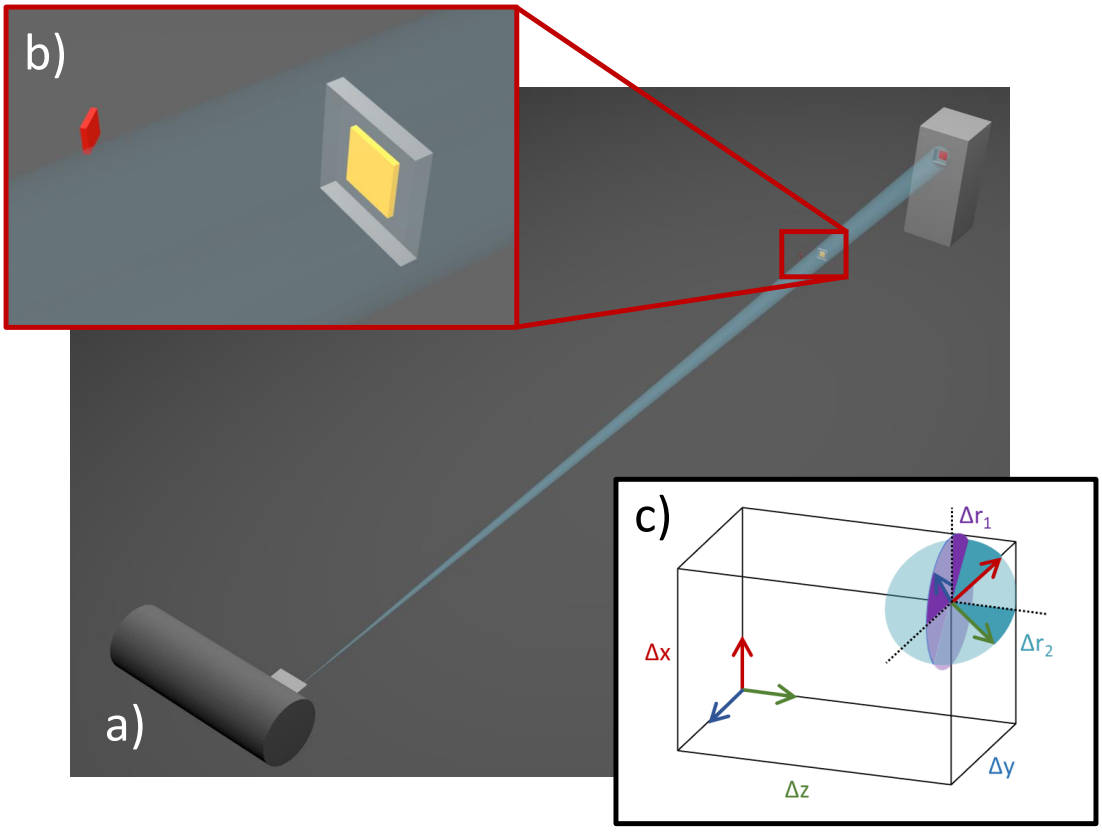

Figure 3.4: The overall simulation geometry is depicted in a) and the target powder and XRF detector geometries are highlighted in b). Geometric random sampling is used on the placement of the target powder in five dimensions, shown in c), Δx, Δy, Δz, Δr_1, and Δr_2.

ometry are randomly sampled. The values are limited such that the powder remains in view of the imaging detector and the maximum rotation is wider than observed in experiment, visualized in Figure 3.4c. Three translational and two rotational dimensions are randomly sampled at the start of each of these simulations. These randomizations were designed to cover a wide range of experimental uncertainties. The purpose of this dataset was to train a deep learning model to perform quantitative material discrimination with geometric uncertainty which would relax the operational requirements of the system.

In some cases, the target was filled with a single powder compound such as UO_2, Bi_2O_3, or W. In other cases, the target may be a mixture of multiple powders. The total bulk density d_{mixed} of the mixed powder is calculated in Equation 3.9 from the mass fractions f_i and the

constituent bulk densities d_i.

$$d_{\text{mixed}} = \prod_i d_i \Big/ \sum_i \left[f_i \prod_{j \neq i} d_j \right] \tag{3.9}$$

The transmission detector collects the flux 205 cm from the source after it the particles have traversed the geometry. The sensitive area is 2 cm $\times$ 2 cm with 6400 pixels that are 250 μm $\times$ 250 μm, the dimensions of a HEXITEC detector [64]. The XRF detector in the simulation collects the flux incident on a 7-mm diameter disk representing the Be window of the detector. It is located 3 cm to the left and 3 cm toward the source from the target center. The radiation interactions inside the detectors were not modeled in these simulations. Instead, they tallied flux distributions which were transformed as described in Subsection 3.1.2. Simplifying the detectors in this way is an efficient use of computational resources.

3.2.3 Generating Spectral Radiography and X-ray Fluorescence Data

A point source generated two particle types simultaneously: X-rays and geantinos. The geantino is a fictional particle type created for Geant4 simulations. They do not interact with the materials but they are transported through the geometry, making them useful for gathering information about the simulation. All particles in the simulation were emitted in the direction of the target within a small solid angle. An additional software package, SpekCalc, was used to generate the X-ray source's energy distribution that matched a 160 kV Comet XRS-160 X-ray source, $\Phi_0(E)$ [82], shown in Figure 3.5. The X-ray energies were sampled from this distribution at the start of each event.

The X-rays proceed with a traditional Monte Carlo simulation, traversing the simulation's world by randomly sampling from interaction probabilities. The geantinos traverse the geometry without interaction. As a geantino steps through each volume along its path, it accumulates the pathlength of each material it encounters.

The transmission detector is only sensitive to the geantinos. When a geantino hits a

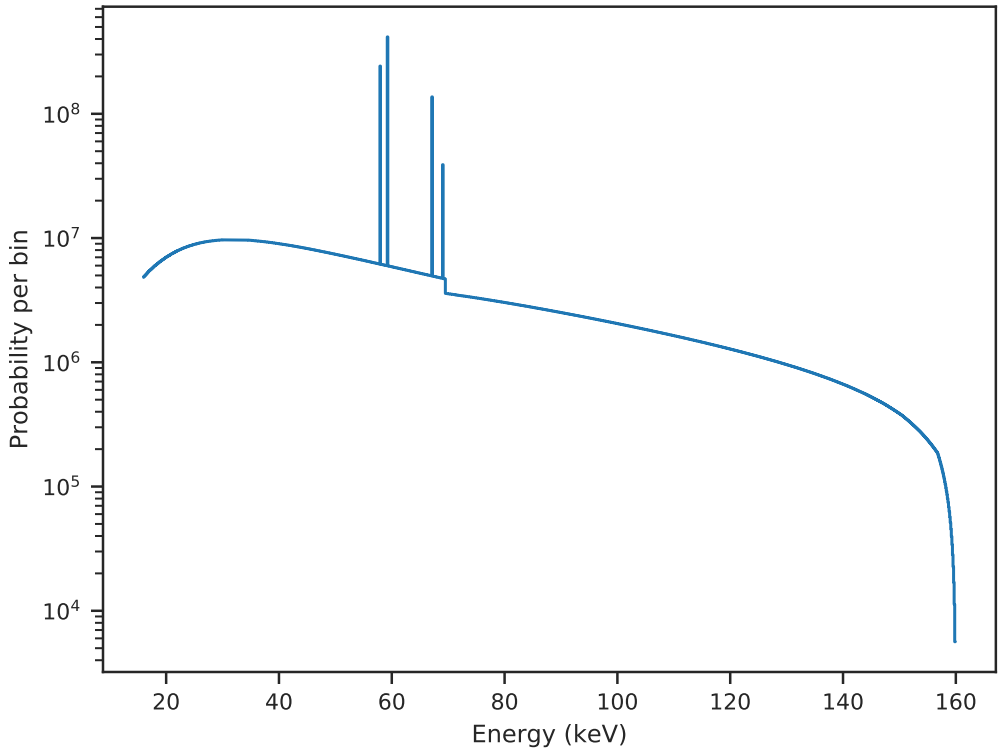

Figure 3.5: The source distribution $\Phi_0(E)$ used in simulations was generated using SpekCalc.

pixel, the X-ray energy spectrum $\Phi(E; \boldsymbol{\rho})$ is calculated with Equation 3.1 using the accumulated areal densities, $\boldsymbol{\rho}$, and the source distribution. The final spectrum of a pixel is the average over the spectra resulting from all the incident geantinos. A pixel's total intensity is kept proportional to the number of geantino hits. This approach is spatially stochastic and spectrally deterministic and reduces the number of required source particles from $>$ 10^{11} to 10^7.

The XRF detector is sensitive only to incoming X-rays which generally result from X-ray fluorescence or scatter. The source X-rays are used to generate the emissions within the sample that are measured by this detector. These detection events are relatively low probability, so we used 10^9 source X-rays per simulation.

3.2.4 Extracting Pixel-Wise Elemental Mass

An important feature of a simulated environment is perfect system knowledge. The elemental masses projected onto each pixel were used for both testing the material discrimination methods and training deep learning models. The randomized geometry and finely voxelized powder target made it difficult to determine the pixel-wise, elemental masses. In addition to the spectral transmission, the geantinos were used to determine these masses.

The projected mass on a pixel k is the total mass within the pyramid whose base is the pixel area and apex is the source origin. We first define $\hat{y}$ as the vector pointing from the source to the center of the detector face. The momentum geantino traveling toward the detector is primarily in the $\hat{y}$ direction. A step within a uniform material volume has a starting point y_{pre} and ending point y_{post} in the $\hat{y}$ direction. The volume within the pyramid bounded by y_{pre} and y_{post} is temporarily assumed to be filled with the same density. Using the material's density d, the pixel width $w = 250$ μm, the source-to-detector distance $h = 205$ cm, we approximate the elemental mass for element i with weight percent $(\text{wt}\%)_i$ as

$$m_i = (\text{wt}\%)_i * \frac{d}{3} \left(\frac{w}{h}\right)^2 \left({y_{\text{pre}}}^3 - {y_{\text{post}}}^3\right). \tag{3.10}$$

The m_i values for an element are summed along the entire path of the geantino and accumulated for each pixel such that each geantino estimates the total elemental masses viewed by a pixel. The final mass is the average of these estimates.

3.3 Experimental Materials and Methods

3.3.1 Equipment and Materials

The experimental setup consisted of three components: the X-ray source, powder targets, and the detector. The source was a 160 kV Comet XRS-160 X-ray generator. The detector was a HEXITEC pixelated, spectroscopic X-ray detector with a 1-mm CdTe detection vol-

ume. The targets were powder volumes that were sealed in 3D printed cuvettes. We used used Bi_2O_3 in the experiments. The cuvettes were 3D printed using a Monoprice MP Mini SLA LCD High Resolution Resin 3D Printer. The printer used MP Rapid UV 3D Printer Resin, an ultraviolet curing epoxy.

All powder material was used as received. The bismuth(III) oxide MicroPowder (Bi_2O_3, 99.9% trace metal basis, 1.05 g/cm_3 bulk density) was obtained from US Research Nanomaterials, Inc.

3.3.2 3D Printing Cuvettes

The 3D printed cuvette design comprised a base component and a lid component. A 3D model was first designed using Blender, depicted in Figure 3.6. The base had a well for powder that is 1 cm $\times$ 1 cm $\times$ 1 mm, and the lid is designed to fit over the well once filled. The lid required 0.25 mm margins to comfortably fit over the base.

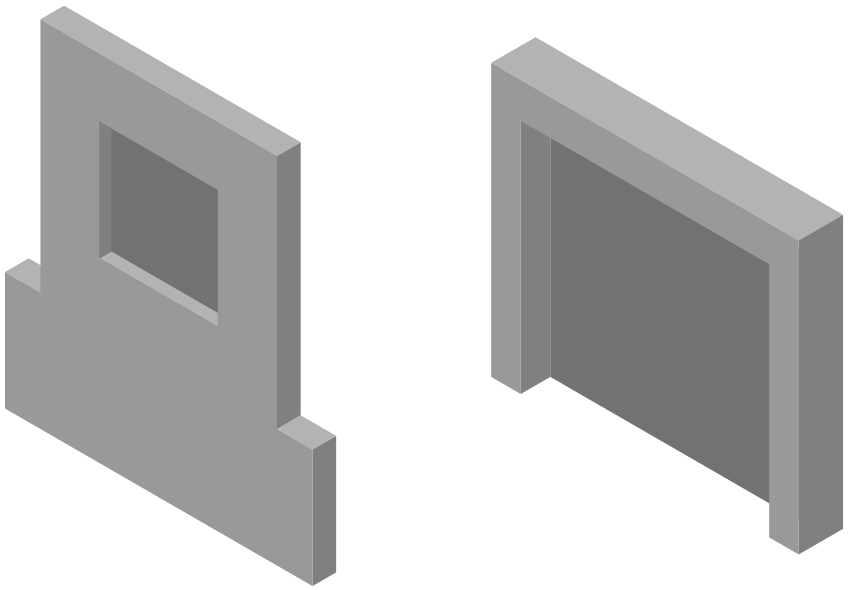

Figure 3.6: The cuvette consisted of two 3D printed pieces. The base (left) features a well for powder. The lid (right), combined with Kapton tape and UV resin, sealed the powder volume.

Using ChiTuBox, the supports and bases necessary for the 3D printing process were

generated for the design. The supports are essential to a successful resin 3D print by securing the print to the build plate and securing areas prone to deformation. A single print contained multiple bases and lids with added supports. The result was then sliced into a file made up of printable layers and printing instructions.

The printer exposed 50 μm slices at a time to a 405 nm UV light to be cured. The print cross sections were formed by masking the UV source with an LCD screen. The first 3 layers were exposed for 30 s each to ensure adhesion to the build plate. Subsequent layers were cured for 8 s each. After each exposure, the cured layer was pulled away from the bottom of the vat to make room for the next layer. Debris from previous print jobs, insufficient supports, isopropynol-contaminated resin, and cold resin were leading causes of failed prints.

After printing, the prints were cleaned with isopropyl alcohol and water. We then cured the prints under a UV lamp for another 30 min. Once hardened and no longer tacky, the supports were cut away, leaving large imperfections. Additional surface imperfections in the form of small, parallel ridges are caused by the discrete layers and limited precision. While these artifacts have negligible impact on the X-ray signatures, rough surfaces interfere during sample preparation. The thicknesses were marginally increased to permit sufficient sanding. The dimensions of the base and lid including all margins used during printing are detailed in Figure 3.7 and Figure 3.8, respectively.

Once the supports were removed, the the faces were sanded down to the target thicknesses: 2 mm for the base and 1 mm for the lid. A micrometer was used to measure thickness. The sanding was done in four stages, starting with 220 grit sand paper and increasing to 3000 grit. The faces were then smooth enough to easily clean off excess powder.

3.3.3 Powder Sample Preparation

Each cuvette was filled with Bi_2O_3 and sealed with the following procedure. First, the base and a piece of 12-mm Kapton tape were weighed. Next, the powder was placed and pressed

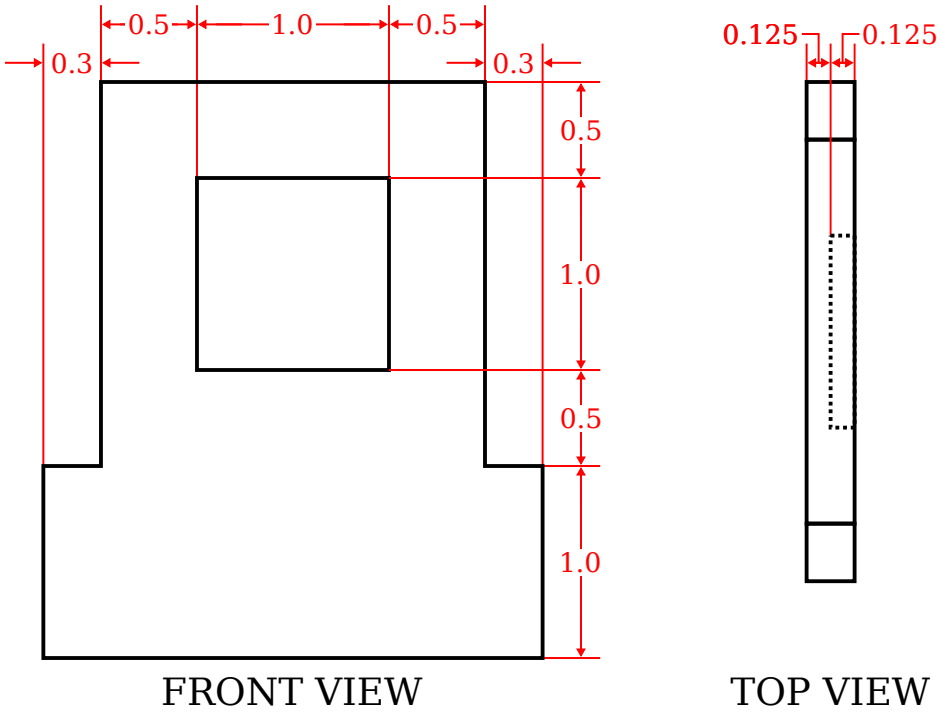

Figure 3.7: The dimensions of the base of the cuvette that were used by the printer had a thickness margin to accommodate sanding. All values are reported in cm.

into the well of the base using a spatula. Once full, the powder was leveled with a straight edge and the excess powder was cleaned from the cuvette, Figure 3.9. Next, the piece of Kapton tape was placed over the powder volume and weighed again, Figure 3.10. Next, a small amount of UV curing epoxy was brushed into the inside of the lid which was then pressed over the Kapton tape. The sample was then cured under a UV lamp for at least 24 hours, Figure 3.11.

3.3.4 Procedure for Experimental Spectral Radiography of Powders

The spectral radiography measurements shared a system geometry with the simulations described in Subsection 3.2.2. The source and detector were mounted and aligned using a laser level, Figure 3.12. The pinhole 2-mm W collimator was then aligned using the X-ray source and HEXITEC radiographs. The collimator was horizontally translated in one direction until only half the image was exposed; this was repeated in the opposite direction,

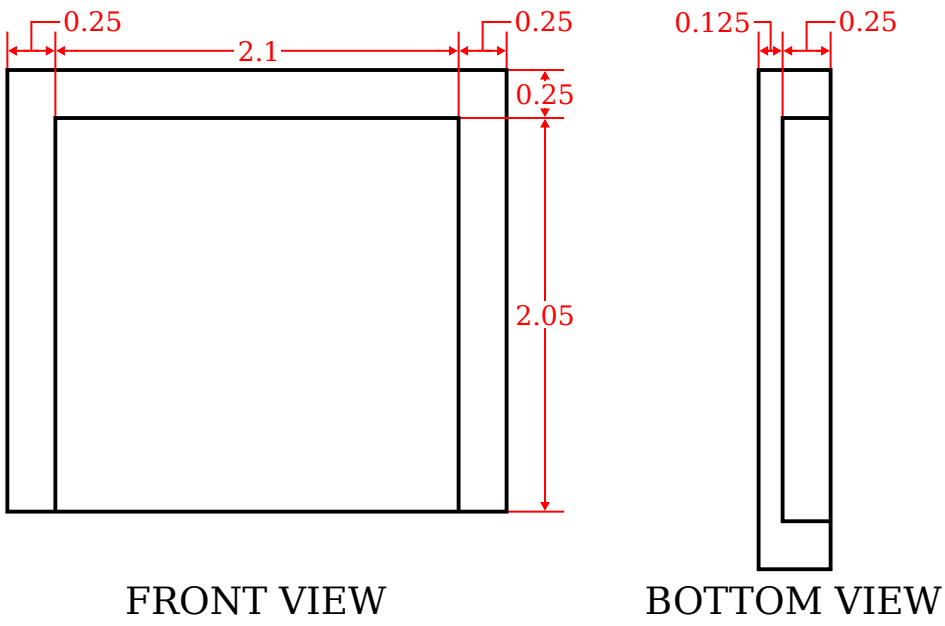

Figure 3.8: The lid of the cuvette had geometric margins to fit over the base and a small thickness margin to accommodate sanding. All values are reported in cm.

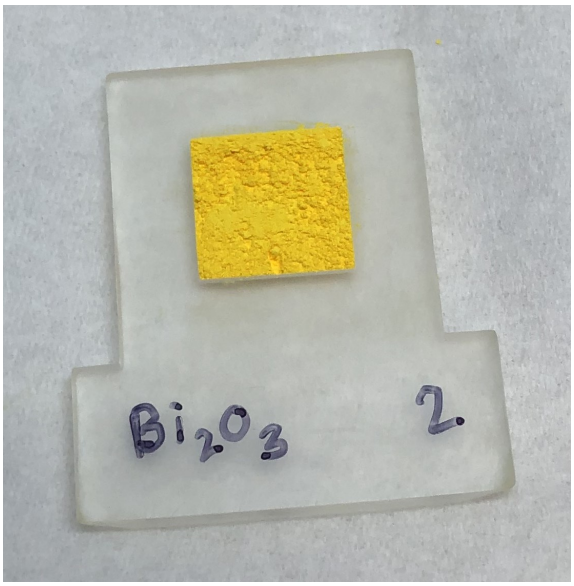

Figure 3.9: The cuvette is filled with powder (Bi_2O_3 shown here) which is pressed and leveled off.

and the centered position was taken as the average. The vertical alignment was performed in the same way.

The X-ray generator was operated at 160 kVp and 40 μA with a 0.4-mm focal spot.

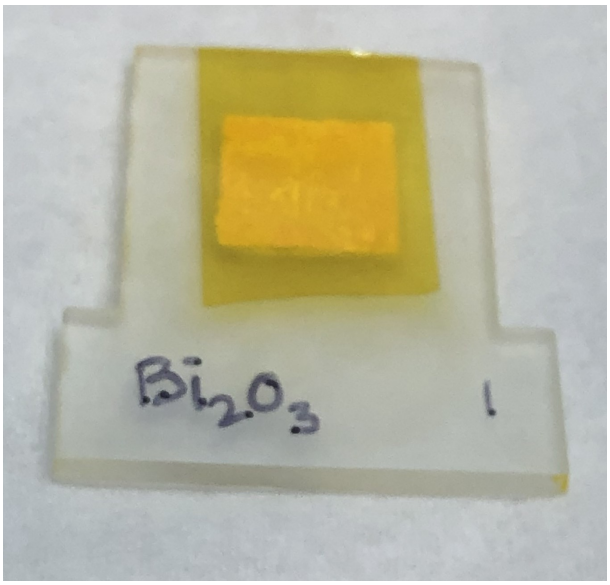

Figure 3.10: Kapton tape is placed over the powder volume as an initial seal. It prevents powder from spilling during the second weighing and provides a barrier between the powder and epoxy.

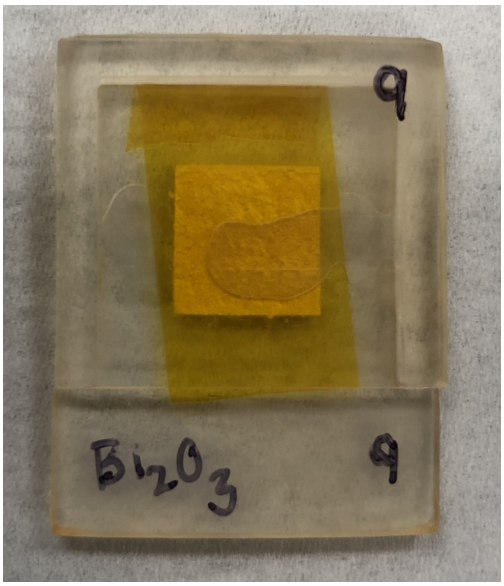

Figure 3.11: A finished sample has been sealed with UV curing epoxy to prevent powder leakage.

The HEXITEC detector was biased to -500 V and output frames at 8 kHz. Bias refresh was used to mitigate the polarization due to space charge buildup in CdTe [52]. In bias refresh, the bias is turned on and acquires data for a set duration t_{acq}, turned off for a duration t_{off}, turned on without acquiring for a duration t_{holdoff}, and then acquisition recommences.

Figure 3.12: The source and W pinhole collimator (bottom left) are aligned with the detector (top right) in two steps: approximate alignment with laser level and finding pinhole center with X-ray source and detector.

The powder samples were held by a mount fixed to the optical table 57 cm from the detector face. A laser level first was aligned with the centers of the detector and source. The cuvette was then placed in the mount, approximately centered with the level, Figure 3.13. An initial X-ray image was taken to ensure the powder volume was fully viewed by the detector.

The acquisitions had live times of 10 min resulting in $\approx$ 61.6 GB of data in the form of frames. The flatfield was measured on multiple days at various times of the day. Each sample was measured up to 6 times, and the sample placement direction was alternated.

3.3.5 Parsing Experimental Data

The data files from a HEXITEC acquisition are large binary files made of linearized frames. There are 6400 pixels per frame, and each pixel has a 2 byte value representing the size of the charge pulse measured in that pixel during that frame. In order to read, process, and reformat the data into histograms, we used C++ over Python for its speed.

The reading was optimized by implementing a producer-consumer model for thread

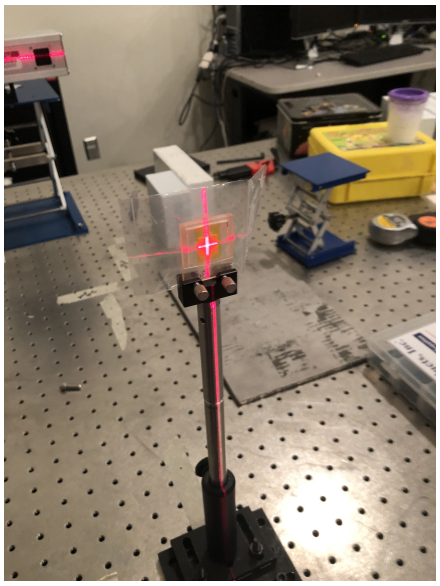

Figure 3.13: The cuvette centered was done using a laser level aligned with the source and detector.

management. The producer thread read in binary frames as arrays of integers then added each frame to a queue. Only one reader was used because the hard drive reads fastest sequentially without jumping around in the file. Additionally, all the data was stored on an external hard drive.

A consumer thread was responsible for parsing the frames. It started by retrieving a frame from the queue or waiting when empty. The first processing step was to apply the CSD discussed in [54] to correct for charge sharing. Charge sharing occurs in small pixel detectors when the induced charge cloud diffuses and is measured by adjacent pixels [83]. After CSD removes events from the frame, the remaining values are calibrated using the pixel-wise energy calibration presented in [77]. A 3D histogram is used to accumulate the resulting calibrated frame data above 5 keV. The histogram has dimensions $80 \times 80 \times 600$ such that the 600 energy bins range from 0 to 170 keV. Saving the data in this way reduces the 61.6 GB file to around 6 MB, a factor of 10^4.

3.3.6 Flatfield Acquisition and Scaling

The flatfield measurement is the detector's response to the beam in the absence of a target. The analysis of attenuation inspects the spectral change relative to the flatfield. In the pixelated HEXITEC detector, the flatfield is useful in normalizing pixel-to-pixel variations in efficiency. These variations are visible in the energy-integrated flatfield images in Figure 3.14. The majority of pixel efficiencies appeared to vary about some mean efficiency; in contrast, the images contained a small number of dead pixels with few to no counts, likely due to a poor bump bond. The pixels with a number of counts higher than the majority of pixels were located next to dead pixels and along some edges. The elevated counts were artifacts from the CSD algorithm (Subsection 2.2.5) which attempted to remove charge-sharing events. Fewer counts were removed in these pixels because they were not surrounded on all sides by functioning pixels.

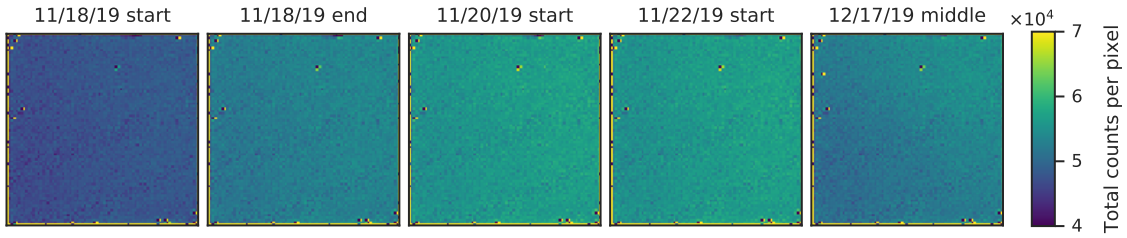

Figure 3.14: Energy-integrated images of the five 10-min flatfield measurements. While the images share spatial features and artifacts, the total number of counts is inconsistent across measurements.

The total count rate also fluctuated. For the same 10-min acquisition time, the variations in count rates were observed across different days and different times within the same day. Other variations were caused by different bias refresh settings and dropped frames during acquisition. Figure 3.15 shows the number of counts per frame over 10 min from five flatfield measurements. The CSD correction was applied to all measurements. The baseline count rate for each measurement varied. For example, one end-of-the-day measurement (orange) was significantly lower than the others. The blue line does not contain the same

total number of frames, indicating a block of frames was dropped. The purple line has spikes because it was measured with a shorter bias holdoff time. Before CSD, this purple acquisition had fewer counts per frame than most others and subsequently had the lowest CSD removal fraction. A lower removal fraction suggests less charge sharing and detector polarization [84].

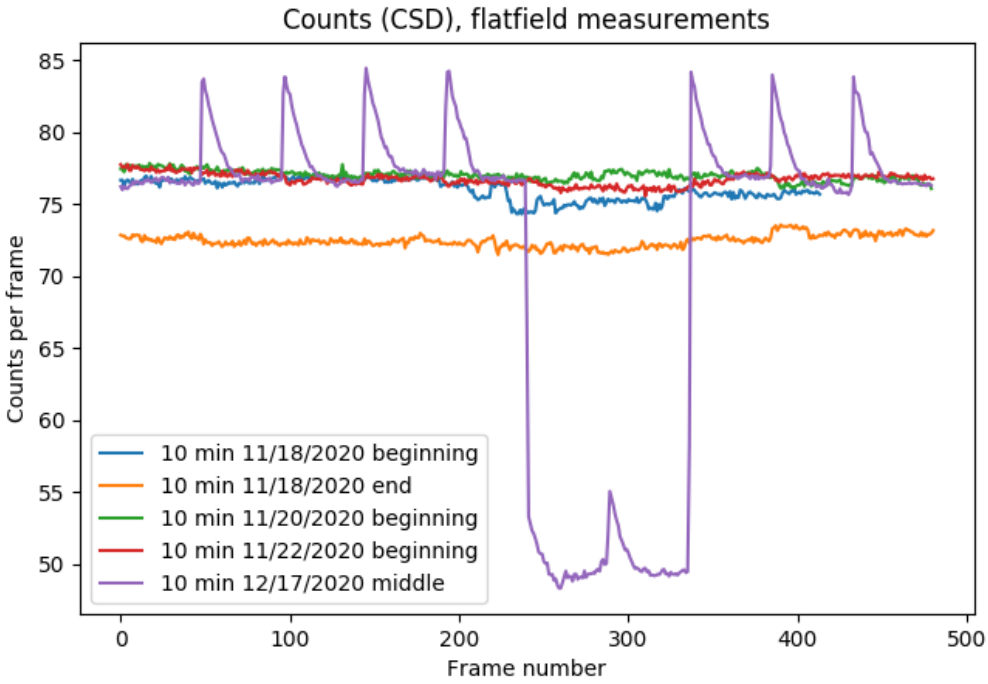

Figure 3.15: The total counts per frame of the five 10-min flatfield measurements show different behavior. The measurement at the end of the day (orange) showed the lowest efficiency. The measurement with jumps in count rate (purple) had a shorter bias holdoff time.

The variations in spectral response were also considered. The spectra summed over all pixels for each measurement are shown on the left in Figure 3.16. The spectral differences can be partially explained by the difference in total efficiency which can be corrected by normalization. Polarization in the detector can also cause spectral differences because it reduces the electric potential and shrinks the depleted region [53]. Higher energy events result in larger charge clouds in the detector. In the case of a lower electric potential, these larger clouds are more likely to result in charge sharing as they diffuse laterally to nearby

pixels. The CSD algorithm removes events occurring in adjacent pixels within the same frame leading to fewer counts at higher energies. By restricting the spectra to > 50 keV, we determined the normalized spectral, shown on the right in Figure 3.16, were consistent with each other.

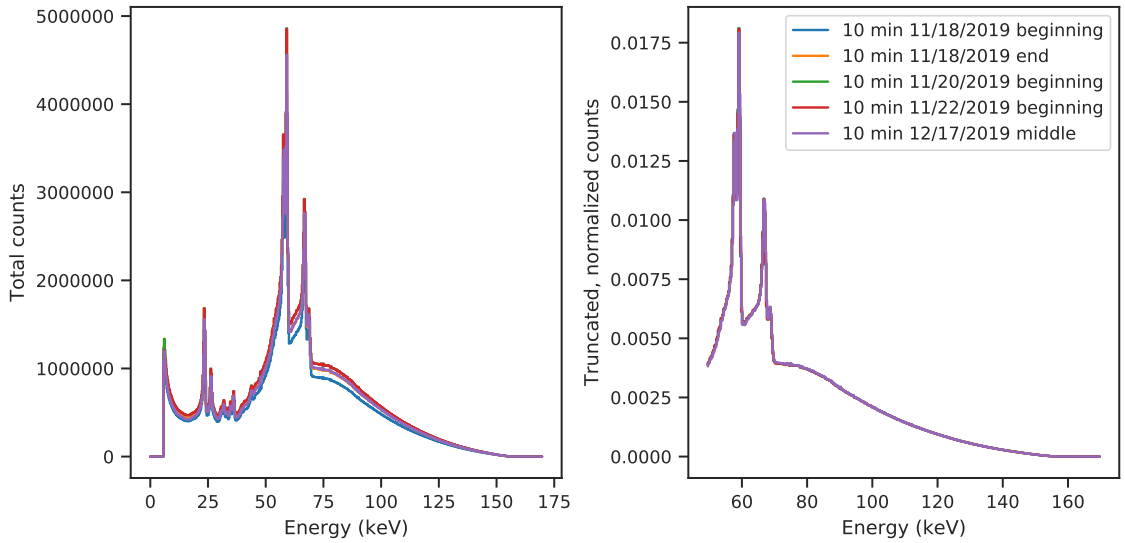

Figure 3.16: The spectra summed over all pixels (left) from five 10-min flatfield measurements show differences in intensity and distribution. After truncating energies at 50 keV then normalizing to number of counts (right), the spectra align with each other.

A total of five 10-min flatfield measurements were acquired on multiple days either at the start, in the middle, or at the end of an all-day series of 10-min powder measurements. We combined these five measurements to form a flatfield distribution with higher confidence. This combined flatfield was used to analyze attenuation across all measurements with targets. The target measurements also suffered from the same variations as the flatfield measurements; therefore, a scaled flatfield was determined and used for each target measurement, outlined in the following steps:

1. Acquire five 10-min flatfield measurements across different days and times.

2. Truncate spectra to only include energies > 50 keV.

3. Sum measurements and normalize by total number of counts.

4. For each target measurement:

   (a) Manually select a bounding box on energy-integrated image around the target area, ensuring to exclude any pixels in view of the target, shown in Figure 3.17

   (b) Sum over all pixels outside bounding box boundaries for target measurement (s_i) and truncated, normalized flatfield measurement (s_0).

   (c) Calculate flatfield scale factor $f_i = s_i/s_0$.

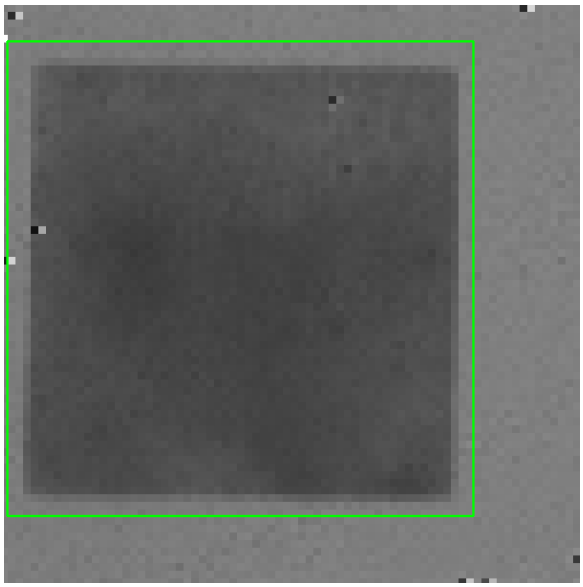

Figure 3.17: Selection of bounding box to exclude target area from flatfield scaling calculation.

The truncated, normalized flatfield was scaled by f_i for the i^{th} target measurement. This procedure has important underlying assumptions, listed below.

Assumption: Flatfield measurements share a data-generating probability distribution within the spectral region of interest.
Assumption: Flatfield and target measurements share an intrinsic pixel-by-pixel detector response.
Assumption: The scattering response and the attenuation through the cuvette are negligible within the spectral region of interest.

CHAPTER 4
A PARTIAL-VOLUME CORRECTION FOR QUANTITATIVE SPECTRAL X-RAY RADIOGRAPHY

4.1 Introduction

Recent years have brought large strides in the development of advanced x-ray detectors [64, 32, 86, 87]. These detectors combine the utilities found across previously distinct technologies, i.e. energy spectroscopy, spatial imaging, and room-temperature operation. With this progress, nondestructive interrogation systems have been developed to utilize the newly available tools to solve challenging problems. Material discrimination is one such task where the goal is to identify or to quantify the materials present in a target.

Spectral x-ray radiography is a material discrimination technique using this technology with potential use in many applications, including international safeguards, medical imaging, and national security. An array of pixels measure the spectral transmission of an x-ray beam through a target material. The additional information about the x-ray energies can be used to discriminate materials or even quantify the amount of each. When the target contains metals or high-Z materials, artifacts can appear in the measurements. A pixel that is only partially covered by a highly attenuating target averages this sharp change in intensity across its face. The measurement is the average response across the pixel, causing partial-volume averaging.

For applications using spectral radiography for quantitative material discrimination, the error introduced by partial-volume averaging is problematic. The error is more prominent in pixelated spectral detectors because the spatial resolution is typically worse than their

charge-integrating counterparts. Most existing work focuses on corrections to traditional computed tomography (CT) reconstruction [88, 89, 9, 90]. The algorithms for spectral x-ray imaging are still emerging, and there is a need for addressing the planar, spectral radiography case.

We previously presented a quantitative material discrimination algorithm using non-linear least-squares [7, 91, 18]. Other methods for material discrimination utilizing spectral imaging include basis material decomposition for materials with low atomic number [21], estimating attenuation profiles with linear combinations of the Legendre polynomials [92], and determining the effective atomic number of a target using a quasi-monoenergetic source [93, 94]. Many other nondestructive testing methods using the detectors have been presented [11, 95–100, 24].

These methods, however, do not address the partial-volume averaging effect. The effect has most prominently been investigated in non-spectral CT methods and positron-emission tomography (PET) [9, 90]. One such method estimated a pixel-wise partial-volume fraction after reconstruction by using material assumptions [88, 89]. Others exploited the additional information from multiple projections, such as the subsampling techniques [101, 102]. The exploitation of spectral features is not a clear extension from these CT applications.

This work modeled a system that employs a HEXITEC CdTe detector for hyperspectral x-ray imaging. Egan *et al.* [12] distinguished two classes of pixelated spectral detectors based on the readout and consequently the spectroscopic capabilities. Multispectral detectors measure single photons as they interact in each pixel. They are typically limited to 5–10 spectral bands and can operate at higher count rates [60–63]. Alternatively, hyperspectral detectors are capable of hundreds of spectral bands and superior energy resolution, but the frame-based readout system limits the maximum count rate [64].

The main purpose of this study was to develop a partial-volume correction to our planar, spectral radiography technique for quantitative material discrimination. The work was guided by the following research objectives:

- To update the physics model from the existing algorithm [7] to account for partial-volume effects.
- To implement the necessary changes in the optimization for the new physics model, including stabilization.
- To compare the updated algorithm to the original algorithm on uranium and oxygen mass prediction from UO_2 samples.

In this work, we first discuss the original implementation of the mass quantification algorithm [7]. We then argue the need for a partial-volume correction and how the physics model is updated to account for this effect. We detail the adaption of the algorithm and present the stabilization techniques used. To test the algorithm's performance, we show the performance on uranium and oxygen mass predictions on UO_2 powders simulated in Geant4. The simulation study provides perfect system knowledge which allows us to isolate the bias due to the physical model used in the algorithms.

4.2 Methods

Our previous work starts with the Beer-Lambert law, the physical basis for traditional material discrimination and material reconstruction [7]. The radiant flux $\Phi_{0,k}(E)$ traveling through M materials toward pixel k is attenuated as

$$\Phi_k(E; \boldsymbol{\rho}_k) = \Phi_{0,k}(E) \exp\left[-(\boldsymbol{\mu}(E))^T \boldsymbol{\rho}_k\right] \tag{4.1}$$

where $\boldsymbol{\mu}(E) : \mathbb{R} \mapsto \mathbb{R}^M$ is the energy-dependent mass attenuation coefficient vector, and $\boldsymbol{\rho}_k \in \mathbb{R}^M$ is the areal density vector for M materials in the projection on pixel k. A component $\mu_m(E)$ has units cm^2/g, and a component $\rho_{m,k}$ has units g/cm^2. The goal of quantitative material discrimination is to solve for the areal density vector $\boldsymbol{\rho}_k$ and mass vector $\boldsymbol{m}_k$ for each pixel.

A pixel's detector response function $r_k(E, E')$ gives the probability of observing energy E' for an incoming photon with energy E [77]. Discretizing the response function over observed energy into L bins gives the response $\boldsymbol{s}_k(E) : \mathbb{R} \mapsto \mathbb{R}^L$. The l^{th} component of $\boldsymbol{s}_k(E)$ is given by

$$s_{k,l}(E) = \int_{E_l}^{E_l+\Delta E_l} r_k(E, E')dE' \tag{4.2}$$

where E_l is the lower-bounding energy of bin l and ΔE_l is the bin width. The spectral image is collapsed to form $\boldsymbol{s}_k(E)$ with $L = 5$ energy bins $E_l \in \{58, 66, 78, 98, 116, 142\}$ in keV. The energies below 58 keV have little transmission signal and primarily result from scatter.

The measured detector response $\boldsymbol{d}_k(\boldsymbol{\rho}_k) \in \mathbb{R}^L$ of pixel k to incident flux $\Phi_k(E; \boldsymbol{\rho}_k)$ is

$$\boldsymbol{d}_k(\boldsymbol{\rho}_k) = \int_0^\infty \boldsymbol{s}_k(E)\Phi_k(E; \boldsymbol{\rho}_k)dE. \tag{4.3}$$

The k^{th} pixel to is described by its row i and column j with response in L energy bins, $\boldsymbol{d}_{i,j}(\boldsymbol{\rho}_{i,j}) \in \mathbb{R}^L$. The responses of all the pixels and L energy bins form a three-dimensional tensor which is vectorized into $\boldsymbol{d}(\boldsymbol{\rho}) \in \mathbb{R}^N$, where $N = 80 \cdot 80 \cdot 5 = 32000$. The term $\boldsymbol{\rho} \in \mathbb{R}^{12800}$ is a similarly vectorized tensor of areal densities of all pixels and of all $M = 2$ materials.

The $\boldsymbol{\rho}$ is determined via least-squares such that $\boldsymbol{d}(\boldsymbol{\rho})$ best approximates the observed detector outputs, $\boldsymbol{d}_{\text{obs}}$. The cost function is defined as

$$F(\boldsymbol{\rho}) = \frac{1}{2}\left\| \frac{\boldsymbol{d}^{(n)}(\boldsymbol{\rho}) - \boldsymbol{d}_{\text{obs}}}{\sqrt{\boldsymbol{d}^{(n-1)}(\boldsymbol{\rho})}} \right\|^2 \tag{4.4}$$

where n is the iteration index.

The task of the algorithm is

$$\min F(\boldsymbol{\rho}) \text{ subject to } \boldsymbol{c}(\boldsymbol{\rho}) \geq 0. \tag{4.5}$$

The components of $\boldsymbol{c}$ are the functions defining the constraints. A nonnegativity constraint on $\boldsymbol{\rho}$ is used, given by $\boldsymbol{c}(\boldsymbol{\rho}) = \boldsymbol{\rho} \geq 0$.

In pixel k, the mass vector $\boldsymbol{m}_k$ is approximated using the average projected area of the target given by

$$\boldsymbol{m}_k \approx \boldsymbol{\rho}_k w^2 \left(1 - \frac{t + 2a}{2h}\right)^2, \tag{4.6}$$

where w is the pixel width, t is the target thickness, a is the target-to-detector distance, and h is the source-to-detector distance. This vector is the mass prediction for each material in each pixel.

4.2.1 Partial-Volume Effect

The physical model in Equation 4.1 has the built-in assumption that the flux passing through a material is uniform across the projected pixel area. This assumption is only valid for small variation in areal density.

Consider a simple case where a single pixel views a material with mass attenuation coefficient μ. A fraction $(1 - f)$ of the flux in the projected area is attenuated by areal density ρ_1, and the remaining fraction f is attenuated by the same material of a different areal density ρ_2, where $f \in [0, 1]$. The true, total areal density ρ_{true} used to determine the mass is

$$\rho_{\text{true}} = (1 - f)\rho_1 + f\rho_2. \tag{4.7}$$

The total flux Φ_{obs} observed by the pixel is the linear combination of the attenuation from each areal density. This is given by

$$\begin{aligned}\Phi_{\text{obs}} =&(1-f)\Phi_0 \exp\left(-\mu\rho_1\right)+\\ &f\Phi_0 \exp\left(-\mu\rho_2\right),\end{aligned} \tag{4.8}$$

where Φ_0 is the unattenuated flux. The algorithm attempts to find some effective areal density ρ_{eff} that satisfies

$$\Phi_{\text{obs}} = \Phi_0 \exp\left(-\mu\rho_{\text{eff}}\right). \tag{4.9}$$

Theorem 1. For some ρ_{eff} used to estimate Φ_{obs} in Equation 4.9 where Φ_{obs} is given by Equation 4.8, $\rho_{\text{true}} \geq \rho_{\text{eff}}$.

The proof for this theorem can be found in Appendix A.

A special case of this theorem occurs at the boundary of the target where $\rho_1 = 0$. This idea is illustrated in Figure 4.1. When $\rho_1 = 0$, we consider the beam unattenuated in that area because no attenuation occurs without target material.

When using the model in Equation 4.1, ρ_{eff} is limited to estimating a uniform density in each pixel resulting in error where there are sharp density gradients. In this work, we propose an updated physical model that accounts for the partial-volume effect where the beam is partially attenuated by target material and partially unattenuated.

4.2.2 Updating the Physics Model

We have updated the physics model with an additional fraction parameter f_k. This new parameter represents the fraction of the area of pixel k that sees flux attenuated by $\boldsymbol{\rho}_k$ such that the remaining fraction $(1 - f_k)$ sees an unattenuated flux. This updated physics model is

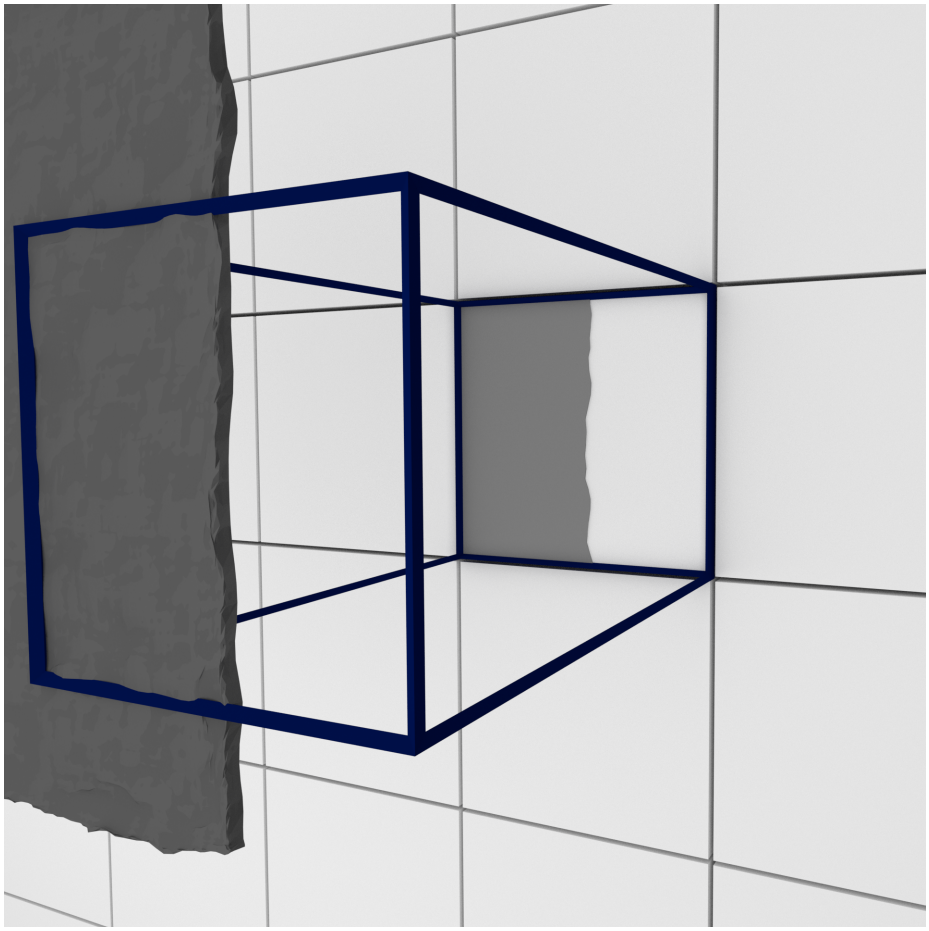

Figure 4.1: Partial-volume attenuation in a pixel occurs when a fraction f of the area viewed by a pixel is sample and the remaining fraction $(1-f)$ is unattenuated. This visual representation shows a pixel viewing the edge of the sample volume. The cross-sectional area of the portion of the sample in view of the pixel is fA where A is the total area in view of the pixel.

$$\begin{aligned}\Phi_k(E,\boldsymbol{\rho}_k) = \; & (1-f_k)\Phi_{0,k}(E) \\ & + f_k\Phi_{0,k}(E)\exp\left[-(\boldsymbol{\mu}(E))^T\boldsymbol{\rho}_k\right].\end{aligned} \tag{4.10}$$

We make a physical argument that the fraction f_k must be constrained to $0 \leq f_k \leq 1$. The original algorithm [7] performs a gradient projection step, then a Gauss-Newton step with a reduced Hessian. Both steps are subject to a nonnegativity constraint on ρ as described in [103, p. 157-161]. The new variable f_k requires this same nonnegativity

constraint with an additional upper bound constraint at $f_k \leq 1$. We define the fraction as $f_k(x_k; \varepsilon)$, a function of a variable x_k with a scaling parameter ε. Different functional forms, e.g. linear and exponential, and scaling values were explored for stability and performance. The f_k and corresponding x_k values in all pixels comprise the vectors $\boldsymbol{f}$ and $\boldsymbol{x}$. The task of the updated algorithm is

$$\min F\left(\begin{bmatrix}\boldsymbol{\rho}\\ \boldsymbol{x}\end{bmatrix}\right) \text{ subject to } \boldsymbol{c}\left(\begin{bmatrix}\boldsymbol{\rho}\\ \boldsymbol{x}\end{bmatrix}\right) \geq 0. \tag{4.11}$$

The new parameters $\boldsymbol{x}$ are now concatenated to the areal density vector $\boldsymbol{\rho}$. To allow for flexibility, such as imposing multiple constraints on a variable, the nonnegativity constraint $\boldsymbol{\rho} \geq 0$ is extended to a more general inequality constraint such that

$$\boldsymbol{c}\left(\begin{bmatrix}\boldsymbol{\rho}\\ \boldsymbol{x}\end{bmatrix}\right) = \begin{bmatrix}\boldsymbol{\rho}\\ \boldsymbol{f}(\boldsymbol{x};\varepsilon)\\ 1-\boldsymbol{f}(\boldsymbol{x};\varepsilon)\end{bmatrix} \geq 0. \tag{4.12}$$

4.2.3 Adapting the Algorithm

The first step of the algorithm is a gradient projection updated using methods described in [104]. To achieve the nonnegativity constraint in this stage, a step vector $\boldsymbol{p}$ is calculated as the negative gradient

$$\boldsymbol{p} = -\nabla F(\boldsymbol{\rho}) = -\boldsymbol{J}^T\boldsymbol{r}, \tag{4.13}$$

where $\boldsymbol{r}$ is the normed residual from Equation 4.4, and $\boldsymbol{J}$ is the Jacobian of $\boldsymbol{r}$. Then $\boldsymbol{p}$ is constrained such that p_i is zeroed when ρ_i is at the boundary and the gradient in that

direction is positive,

$$p_i = \begin{cases} 0 & \rho_i = 0 \text{ and } p_i > 0 \\ p_i & \text{otherwise.} \end{cases} \tag{4.14}$$

This step is scaled by an update parameter $\gamma_{GP}^{(n)}$ and used to update $\boldsymbol{\rho}$ in iteration n. The update parameter is determined from a projected line search such that

$$\gamma_{GP}^{(n)} = \underset{\gamma>0}{\arg\min}\, F\left(\boldsymbol{P}\left(\boldsymbol{\rho}^{(n)} + \gamma \boldsymbol{p}^{(n)}\right)\right), \tag{4.15}$$

where the i^{th} component of projection $\boldsymbol{P}(\boldsymbol{\rho})$ is

$$[\boldsymbol{P}(\boldsymbol{\rho})]_i = \begin{cases} \rho_i & \rho_i \geq 0 \\ 0 & \rho_i < 0. \end{cases} \tag{4.16}$$

Extending this idea to the updated model with new inequality constraints, $\boldsymbol{c}\left(\left[\begin{smallmatrix}\boldsymbol{\rho}\\ \boldsymbol{x}\end{smallmatrix}\right]\right) \geq 0$, we first identify the active set of constraints. This set comprises the constraints where the equality holds, active set $\mathcal{A}\left(\left[\begin{smallmatrix}\boldsymbol{\rho}\\ \boldsymbol{x}\end{smallmatrix}\right]\right) = \{i \mid c_i(\left[\begin{smallmatrix}\boldsymbol{\rho}\\ \boldsymbol{x}\end{smallmatrix}\right]) = 0\}$. We then apply an equality constraint on this set.

For the gradient projection step, the step $\boldsymbol{p}$ satisfies $\nabla c_i(\left[\begin{smallmatrix}\boldsymbol{\rho}\\ \boldsymbol{x}\end{smallmatrix}\right])^T \boldsymbol{p} + c_i(\left[\begin{smallmatrix}\boldsymbol{\rho}\\ \boldsymbol{x}\end{smallmatrix}\right]) = 0$. We define the Jacobian of the active constraints $\boldsymbol{A}^T = [\nabla c_i(\left[\begin{smallmatrix}\boldsymbol{\rho}\\ \boldsymbol{x}\end{smallmatrix}\right])]_{i \in \mathcal{A}(\left[\begin{smallmatrix}\boldsymbol{\rho}\\ \boldsymbol{x}\end{smallmatrix}\right])}$. The step in the gradient projection stage is now

$$\begin{aligned} \boldsymbol{p} = \begin{bmatrix} \boldsymbol{p}_{\boldsymbol{\rho}} \\ \boldsymbol{p}_{\boldsymbol{x}} \end{bmatrix} = \\ -[\boldsymbol{I} - \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}]\nabla F\left(\begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{x} \end{bmatrix}\right). \end{aligned} \tag{4.17}$$

The new projection matrix $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}$ is also used in the projected line search for $\gamma_{GP}^{(n)}$ as in Equation 4.15. It is the orthogonal projector onto the null space of $\boldsymbol{A}$.

The second stage of the original algorithm [7] is a Gauss-Newton step with a reduced Hessian, referenced from [103, p. 157-161]. This starts with calculating gradient $\boldsymbol{g}$, similar to Equation 4.13 where

$$\boldsymbol{g} = \nabla F(\boldsymbol{\rho}), \tag{4.18}$$

and the Hessian matrix $\boldsymbol{H}$ from the Jacobian using the Gauss-Newton approximation

$$\boldsymbol{H} \approx \boldsymbol{J}^T\boldsymbol{J}. \tag{4.19}$$

As described in [103, p. 158-159], a nonnegativity constraint is applied by first finding the active set $\mathcal{A}(\boldsymbol{\rho}) = \{i \mid \rho_i = 0\}$. Using the matrices $\boldsymbol{D_A}$ whose diagonals are 1 for index $i \in \mathcal{A}$ and 0 otherwise and $\boldsymbol{D_I} = \boldsymbol{I} - \boldsymbol{D_A}$, the reduced Hessian and gradient are $\boldsymbol{H_R} = \boldsymbol{D_I}\boldsymbol{H}\boldsymbol{D_I} + \boldsymbol{D_A}$ and $\boldsymbol{g_R} = \boldsymbol{D_I}\boldsymbol{g}$, respectively. The step vector is then calculated as

$$\boldsymbol{p} = -\boldsymbol{H_R}^{-1}\boldsymbol{g_R}. \tag{4.20}$$

Recall that this is a special case of $\boldsymbol{c}(\boldsymbol{\rho}) \geq 0$ when $\boldsymbol{c}(\boldsymbol{\rho}) = \boldsymbol{\rho}$. To extend the algorithm's capability, we consider the more general approach from [105, p. 457 - 459] to inequality constraints by again calculating the Jacobian of the active constraints $\boldsymbol{A}$. For matrix $\boldsymbol{Y}$ that spans the range of $\boldsymbol{A}$ and matrix $\boldsymbol{Z}$ that spans the null space of $\boldsymbol{A}$, a step is broken into

$$\boldsymbol{p} = \boldsymbol{Y}\boldsymbol{p_Y} + \boldsymbol{Z}\boldsymbol{p_Z}. \tag{4.21}$$

The components $\boldsymbol{p_Y}$ and $\boldsymbol{p_Z}$ are determined by solving

$$
\begin{aligned}
(\boldsymbol{AY})\boldsymbol{p_Y} &= -\boldsymbol{c} \\
(\boldsymbol{Z}^T\boldsymbol{HZ})\boldsymbol{p_Z} &= -\boldsymbol{Z}^T\boldsymbol{g},
\end{aligned}
\tag{4.22}
$$

where $\boldsymbol{Z}^T\boldsymbol{HZ} = \boldsymbol{H_R}$ is the reduced Hessian and $\boldsymbol{Z}^T\boldsymbol{g} = \boldsymbol{g_R}$ is the reduced gradient. The update from the Gauss-Newton stage is then determined as before in Equation 4.15 with a line search for $\gamma_{GN}^{(n)}$.

Finally, the previous algorithm iterations terminated when the Gauss-Newton step size fell below a certain size. The termination condition is replaced by first calculating a mass vector $\boldsymbol{m}$ from the outputs $\boldsymbol{\rho}$ and $\boldsymbol{f}$. In pixel k, the updated mass vector $\boldsymbol{m}_k$ is

$$
\boldsymbol{m}_k \approx f_k \boldsymbol{\rho}_k w^2 \left(1 - \frac{t + 2a}{2h}\right)^2. \tag{4.23}
$$

This vector is the mass prediction for each material in each pixel. The algorithm stops when $\|\Delta \boldsymbol{m}\| \leq 10^{-6}$ where $\Delta \boldsymbol{m} = \boldsymbol{m}^{(n-1)} - \boldsymbol{m}^{(n)}$ is the difference in mass vectors from the previous iteration to the current.

4.2.4 Functional Forms of f

The choice of f affects the stability and performance of the optimization. As mentioned, it can be used to naturally constrain its range to $[0, 1]$. Additionally, the behavior of f and its derivative both near and far from the solution determine its feasibility.

We explore a number of these functional forms for different reasons. These are the monomials εx, εx^2, and εx^3 and the exponentials $e^{\varepsilon x}$ and $\sigma(\varepsilon x)$. The sigmoid function $\sigma(\varepsilon x)$ asymptotically approaches 0 as $x \to -\infty$ and approaches 1 as $x \to \infty$. These forms are of interest because of well-behaved derivatives, computational expense, similarity to the cost function's dependence on ρ, and/or the natural bounds of the function's range. Much of the performance of these forms depends on the derivatives of Equation 4.10,

$$\begin{aligned}\frac{\partial \Phi_k}{\partial \rho_{k,i}} &= -\mu_i f_k \Phi_{k,0}(E) \exp\left[-(\boldsymbol{\mu}(E))^T \boldsymbol{\rho}_k\right] \\ \frac{\partial \Phi_k}{\partial x_k} &= -\frac{df_k}{dx_k} \Phi_{k,0}(E) \left[1 - \exp\left[-(\boldsymbol{\mu}(E))^T \boldsymbol{\rho}_k\right]\right].\end{aligned} \tag{4.24}$$

Forcing these derivatives to have the same sign ensures that the Hessian in Equation 4.19 stays positive definite.

4.2.5 Stabilization

The step calculation in the Gauss-Newton stage in Equation 4.22 causes instability after the update. This arises because the reduced Hessian is ill-conditioned as it approaches the solution, primarily due to pixels measuring unattenuated flux. Recalling from Equation 4.19 that the Hessian is calculated from the Jacobian, the derivatives of Φ from Equation 4.24 determine the behavior of $\boldsymbol{H}$. As $f \to 0$, so does $\frac{\partial \Phi}{\partial \rho}$, and as $\rho \to 0$, $\frac{\partial \Phi}{\partial x}$ also approaches 0. The steps must be larger and larger to make up for the diminishing Hessian in Equation 4.22.

Fraction and areal density are, in a sense, competing towards zero in unattenuated pixels. When the fraction gets very small, the areal density step becomes large in the Gauss-Newton stage due to the small values in the Hessian. Without an upper bound, ρ can have chaotic behavior and only converges with proper hyperparameters and initial conditions.

We explore four techniques to achieve stability with this algorithm.

1. Setting $\boldsymbol{\rho}_k = 0$ when $f_k = 0$ and setting $f_k = 0$ when $\boldsymbol{\rho}_k = 0$ in pixel k
2. Imposing a minimum fraction value greater than 0 in each pixel such that $f_{min} \leq f \leq 1$, similar to the ϵ-active set in [103, p. 159]
3. Setting a gradient threshold g_{th} such that $g_k = 0$ when $|g_k| < g_{th}$ and placing $\boldsymbol{\rho}_k, f_k$ of that pixel into the active set

4. Choosing fraction functional form $f(x;\varepsilon)$ and adjusting ε values

4.2.6 Geant4 Simulations

The full Monte Carlo method simulates x-rays in Geant4 [80, 81]. The geometry was designed so that the assumptions in Equation 4.1 are valid, namely negligible scattering contribution and a parallel beam. The x-rays are emitted in a 0.5° cone with a 160 kVp energy distribution from SpekCalc [82]. The beam immediately enters a 2 cm thick Al filter to reduce the low-energy flux. A harder energy spectrum is advantageous for probing K-edges. The filter is followed by a 1.25 cm thick W collimator with a 2 mm diameter aperture to remove scattered photons. A layer of lead bricks surrounding the collimated beam provide additional shielding.

The UO_2 target is encased in an acrylic cuvette 197 cm downstream from the source. This distance and collimation ensure a nearly parallel beam during transmission. The 12 mm $\times$ 10 mm $\times$ 1 mm target volume ($t = 1$ mm) has 100×120 voxels, and portion of the voxels at the top are left empty to reflect a partially filled cuvette. The density in each remaining voxel is sampled from a normal distribution, resampling if less than 0.1 g/cm^3. The mean and variance are sampled at the start of the simulation, i.e. $\mu_d = 2$ g/cm^3 and ${\sigma_d}^2 \sim U(4 * 10^{-3}, 4 * 10^{-2})$ (g/cm^3)^2. Examples of energy-integrated spectral radiographs are show in Figure 4.2.

The imaging plane is $h = 205$ cm from the source and comprises an 80 $\times$ 80 grid of $w = 250$ μm CdTe pixels. The detector has an Al casing with a square opening to the active detection volume. The upstream scattering was mitigated via collimation and shielding, but scatter occurs in the cuvette, target, and detector housing which is addressed with a 58 keV energy threshold.

The full simulation using x-rays is computationally expensive. Filling the 6400 energy histograms, each with 160 channels, requires a minimum of 10^{11} source particles. We propose a modified simulation using interaction-less particles called geantinos in a way similar

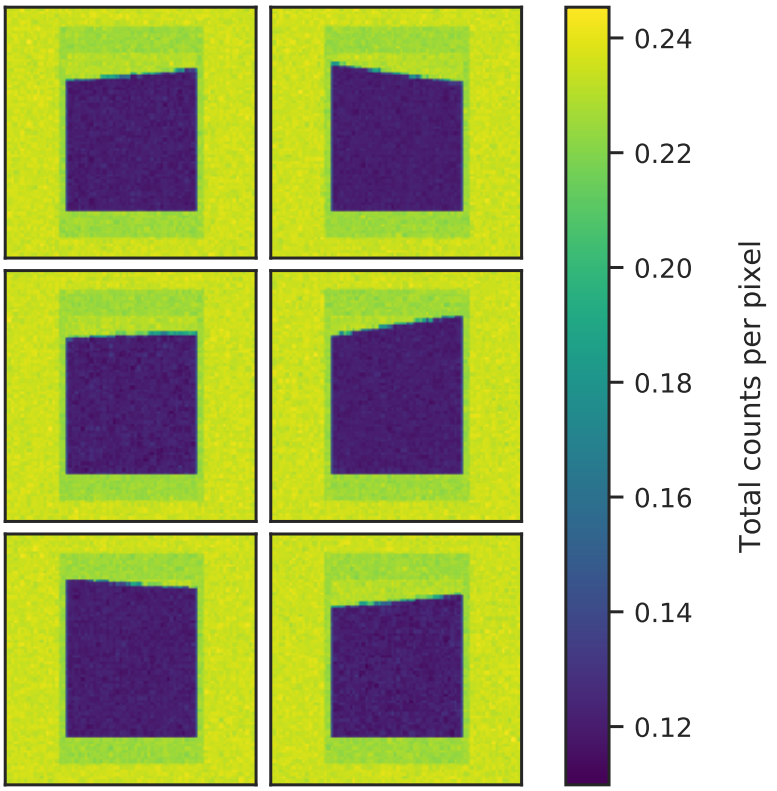

Figure 4.2: The spectral radiographs are generated in the Monte Carlo physics code Geant4. Uranium dioxide powder samples fill acrylic cuvettes up to a random quadratic curve to estimate the irregularity of fill-line. The upper edge accounts for a large portion of the pixels with significant partial attenuation, along with the sample edges. Density variation throughout the powder is achieved through voxelization that is finer than the detector's spatial resolution. These images are energy-integrated spectral radiographs from the Geant4 simulations for visualization.

to MCNP radiography tallies. As a geantino traverse the geometry, it stores material information from its trajectory. When it encounters the imaging plane, the source distribution is transformed using Equation 4.1. Each pixel's response is the average from all geantinos incident on that pixel. The ground truth, elemental mass viewed by each pixel is simultaneously determined. Scaling the output and adding Poisson noise matches the spectral image to the full simulation.

We compared the full and modified simulations to validate the modified approach which does not account for scattered flux. We simulated Φ_0 with an empty cuvette and Φ with a

UO_2 target with uniform density. We considered the energy bins described in Section 4.2. The full simulation of Φ_0 and Φ respectively showed an excess of 3.70% and 3.52% in energies below 58 keV, 1.43% and 1.49% in 58 keV – 66 keV, and $< 1\%$ in higher energy bins. We investigated the significance of this discrepancy by testing the algorithms on both simulation types.

4.3 Results and Discussion

We first compared both algorithms using a single, modified simulation to demonstrate pixel-level performance. The heat maps in Figure 4.3 show the pixel-wise $\Delta m_U = m_U - m_{U,T}$ values where $m_{U,T}$ is the pixel's true projected uranium mass and m_U is the calculated uranium mass. In the original algorithm [7], the error was dominated by pixels at the material boundary and was reduced by an order of magnitude with the updated algorithm. The true and calculated masses are also shown in scatter plots to qualitatively compare distributions. The original algorithm (orange) showed a clear departure from the line of unity at mid-range masses which was absent in the updated algorithm (green). The large negative bias along the boundary is an artifact of partial-volume effect as described in Theorem 1 for the special case of $\rho_1 = 0$. Theorem 1 also explains the negative bias in the rest of the sample where $\rho_1 \neq 0$. Neither algorithm addresses this bias, but this error is minimal compared to the effect along the boundary. The updated algorithm used $f(x) = 100x^2$ and a gradient threshold of 10^{-9} for stabilization.

The update to the algorithm had an even more dramatic shift on the oxygen mass predictions. The heat map from the original algorithm [7] in Figure 4.4 (upper) shows that the oxygen was overestimated in excess of 200 μg along the edges of the material. After the update, the edge values of Δm_O dropped by two orders of magnitude, shown in the lower heat map in Figure 4.4. The scatter plots from the original (orange) and updated (green) algorithms indicate larger deviations from unity on oxygen mass than on uranium mass, relative to their respective masses. Because the oxygen attenuation probability lacks

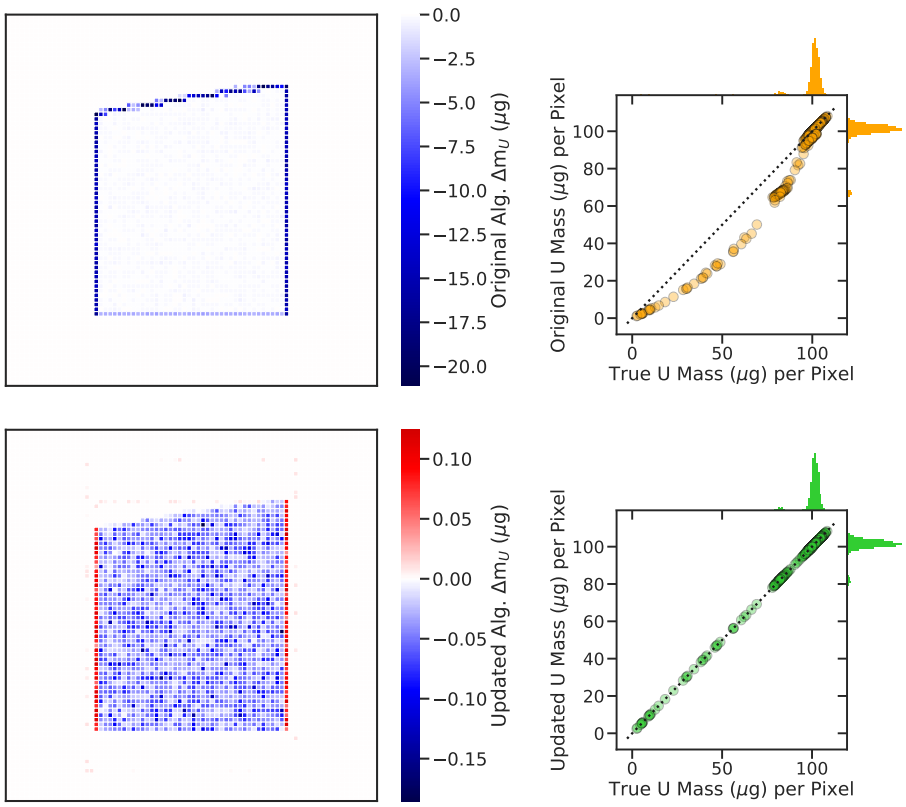

Figure 4.3: The pixel-wise uranium-calculation performances of a single simulation are shown from the original (top row) and updated (bottom row) algorithms. The left column shows images with the difference from ground truth in each pixel (note the difference in scales). Red pixels are overestimated, and blue pixels are underestimated. The right column contains scatter plots of the true mass against the calculated mass in each pixel. The dashed line of unity represents where there is zero error. The original algorithm (orange) tends to underestimate the mass which becomes much less significant with the updated algorithm (green).

spectral features like the k-edge in uranium, the uranium prediction was typically dominant and first to converge. The algorithm then used oxygen to finely tune the fit causing its overestimation.

Across 100 unique, modified simulations, the original algorithm [7] had an average relative error of 1.4% with a standard deviation of 0.20% on the total uranium mass and an average relative error of 122.2% with a standard deviation of 16.1% on the total oxygen mass. Using the updated algorithm with the same parameters mentioned above, it gave an average relative error of 0.068% with a standard deviation of 0.050% on the total uranium

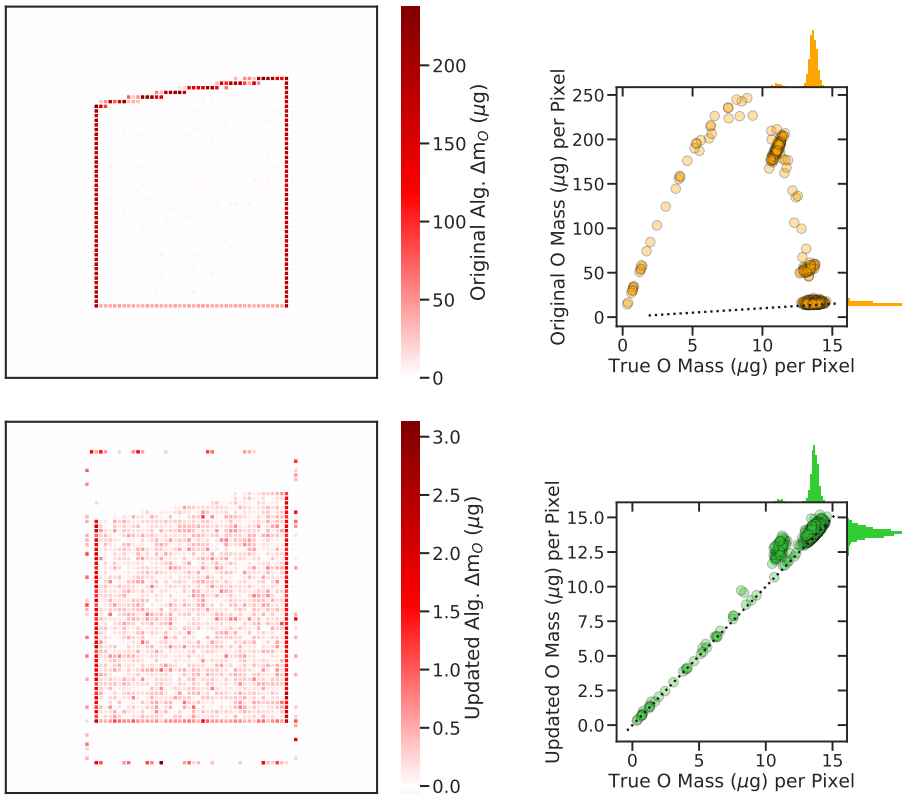

Figure 4.4: Unlike the uranium, the oxygen mass values are typically overestimated (red) by the original (top row) and updated (bottom row) algorithms. The difference from ground truth was reduced by two orders of magnitude in some pixels with the updated algorithm. The largest errors for both cases occurred along the boundary of the material. The right column indicates that the original (orange) and the updated (green) algorithms are less accurate for oxygen than for uranium.

mass and an average relative error of 4.2% with a standard deviation of 2.6% on the total oxygen mass.

Each of the five functional forms and stabilization techniques were also tested on the 100 simulations. The algorithm's performance is a combination of the relative error of total uranium mass, relative error of total oxygen mass, and the number of iterations required to converge. We use a grid search of the following hyperparameters to explore performance.

- $f(x;\varepsilon) \in \{\varepsilon x, \varepsilon x^2, \varepsilon x^3, e^{\varepsilon x}, \sigma(\varepsilon x)\}$
- $\varepsilon \in \{10^0, 10^1, 10^2\}$

- stabilizer $\in \{\text{None}, f_{min}, g_{th}\}$

The values for the stabilizers f_{min} and g_{th} are initially started at 10^{-5} and 10^{-9}, respectively. If any of the 100 examples are unstable with the starting value, it is increased by an order of magnitude repeatedly until stability is reached. Those for which there was no value that stabilized for every example have no recorded relative error or convergence. The results are tabulated in Appendix B.

The first stabilization technique listed, setting $\boldsymbol{\rho}_k = 0$ when $f_k = 0$ and $f_k = 0$ when $\boldsymbol{\rho}_k = 0$, is always applied whenever possible. This includes the results listed as without stabilization. In the case of asymptotic functions, this stabilization is unused unless the f_{min} is also in place. With a $f_{min} > 0$, $\boldsymbol{\rho}_i = 0$ when $f_i = f_{min}$. The initial values are consistent across all results: $f_i(x;\varepsilon) = 0.5$ and $\rho_i = 0.1$.

The gradient threshold stabilization performed best, converging for all fraction functional forms and ε values. The fraction-bounding stabilization also performed well, though it was mostly unstable with $f(x;\varepsilon) = \sigma(\varepsilon x)$, likely due to its diminishing gradients. The algorithm also converged without either of these stabilizers, but this depended on the selection of ε and $f(x;\varepsilon)$.

Lastly, we compared the performances of a modified simulation and a full simulation. The common Φ geometry consisted of a UO_2 sample with uniform density of 3 g/cm^3. The full simulation had 10^{11} source x-rays, and the modified simulation was scaled to the equivalent magnitude. Poisson noise was added to the modified simulation after scaling. We compare the performances using the MSE of the pixel masses in μg, shown in Table 4.1. The updated algorithm used the following hyperparameters: $f(x;\varepsilon) = 100x$, $g_{th} = 10^{-9}$, $f_{min} = 10^{-5}$.

In all simulation cases, the updated algorithm yielded lower MSE for both U and O. The original algorithm [7] showed little difference in performance for different simulation types. The updated algorithm had lowest MSE for the modified simulation with no added noise. The MSE increased by two orders of magnitude for U and three orders of magnitude

Table 4.1: Pixel-wise MSE of mass from different simulation types with common geometry

Sim. type	Original MSE		Updated MSE	
	U	O	U	O
Mod., no noise	23.30	3437.96	0.0025	0.12
Mod., w/ noise	23.65	3591.04	0.48	119.13
Full	22.52	3230.20	0.87	129.42

for O. The full simulation had a much smaller increase in MSE from the modified simulation with noise. The primary difference between the full simulation and the modified simulation with noise is the presence of scattering. This suggests that the updated algorithm is sensitive to noise and has higher estimation error when there is scatter. The increase in MSE from scatter was marginal compared to the increase from noise.

4.4 Conclusion

In this work, we addressed the partial-volume attenuation that occurs in planar x-ray radiography. We extended the algorithm used for quantitative mass discrimination to account for this artifact due to spatial resolution. To do this, the physics model was first updated to include a new parameter f that represents the areal fraction of sample material seen by a pixel. The nonlinear least-squares algorithm with nonnegativity constraints was then extended to use more general constraints. This allowed the fraction parameter to be constrained between 0 and 1.

To test the algorithm's capability, we tested on Monte Carlo x-ray simulations of UO_2 powder generated in Geant4. We explored functional forms for f parameterized by x along with stabilization techniques, showing which hyperparameters perform best. The results show that from simulated data, the updated algorithm improved the calculation of uranium mass from a relative error of 1.4% to 0.068% and of oxygen mass from a relative error of 122.2% to 4.2%.

Work in the lab is ongoing to investigate mass quantification with spectral x-ray detectors. Preliminary results on Bi_2O_3 were presented in [77].

CHAPTER 5
DEEP SPECTRAL X-RAY RADIOGRAPHY

5.1 Introduction

The IAEA has been tasked by the international community to monitor the nuclear activity of non-nuclear-weapon states [1]. The majority of these states also concluded a CSA with the IAEA which allows the IAEA to apply safeguards to facilities that are part of the nuclear fuel cycle. The safeguards measures in place have been designed for material accountancy at a variety of facilities, many of which rely on destructive and non-destructive analysis tools [2]. Some technologies aim to detect partial defects, *i.e.* whether a fraction of the declared amount is missing, and have accuracies on the order of a few percent [2]. Many of these techniques are non-destructive, such as determining uranium enrichment in UF_6 cylinders using γ-ray spectroscopy [106]. The techniques with the highest accuracy ($<$ 1%) typically perform destructive analysis on a sample and are used for detecting bias defects, *i.e.* whether small amounts of material have been diverted time [2]. For example, potentiometric titration is a standard method used to determine the uranium content of non-irradiated material that are on the order of grams [107]. These highly accurate techniques typically require access to a laboratory for analysis [2].

The COMPUCEA technique is an on-site verification assay of uranium concentration and enrichment [108]. The IAEA uses COMPUCEA for routine verification at the Ulba Metallurgical Plant in Kazakhstan, one of the largest nuclear fuel fabrication facilities in the world [5], and it has been widely implemented in European fuel fabrication facilities by Euratom. The technique has been used on solid (powders, pellets, and scraps) and liquid samples. Solid samples must first be dissolved into a nitric acid solution. The solution is passively measured for radioactivity and actively interrogated with X-rays (L-edge densit-

ometry) [108]. COMPUCEA is highly accurate, compact, and transportable. However, the technique also requires laboratory facilities, a multi-step procedure to prepare the solution, multiple 1000 s radiometric acquisitions, and multiple correction factors [5] (*e.g.* for variation in U concentration, the presence of additives like Gd, the thickness of the container's bottom [108]).

Spectral X-ray radiography presents an alternative approach to elemental quantification measurements. Gilbert *et al.* [18] demonstrated the strong potential of spectral X-ray radiography to quantify elemental mass of nuclear fuel powder without chemical preparation. A cornerstone of this technology is the pixelated, spectroscopic X-ray detector which is discussed in detail in Section 2.2. Both spectral X-ray radiography and COMPUCEA measure the spectral X-ray transmission, but the radiographic detector permits spatial variation in concentration because it measures the spectrum incident on each pixel. Compared to the solution samples, the higher density powders also require a higher energy X-ray source in order to be sufficiently penetrating. The K-edge is used as the prominent spectral feature in place of the L-edge at these energies. The CdTe and CZT detectors are efficient in this energy regime, unlike many other semiconductor diode detectors. They also achieve good energy resolution, which is further improved by the use of the small pixel effect [58].

A deep learning formulation of the elemental mass quantification task has a number of advantages. One benefit a deep learning approach could provide is insensitivity to geometric uncertainty. Analytical approaches rely on a well-characterized geometry (*e.g.* sample thickness, location, orientation) explicitly by using geometric parameters in the calculations [18] or geometric correction factors [108] or implicitly with high-precision equipment [34]. By using a training set that incorporates variation in geometric parameters, a deep learning model can learn to predict elemental mass under these conditions.

Deep learning also provides a simple way to combine measurements from different sensors through multimodal fusion [109–112]. The XRF signal that is generated during a transmission measurement is a strong candidate for a second sensor. XRF has been used

to supplement X-ray imaging (non-spectral) by providing elemental identification information [31, 75]. Spectral transmission measurements (non-radiographic) have also been aided by XRF measurements when the sample contains elements at low concentrations. HKED combines a spectral transmission measurement used for uranium concentration with an XRF measurement used for the U/Pu ratio, yielding both the uranium and plutonium concentrations [34]. Erdmann *et al.* [108] proposed supplementing COMPUCEA with an XRF measurement to accurately calculate correction factor for ^{235}U enrichment in Gd-doped samples. XRF signals provide additional information when the concentrations are too low to induce a significant transmission signature.

The aim of this chapter was to determine the viability of deep learning to quantify the elemental mass of powder samples using spectral X-ray radiography and XRF measurements. Realistic simulations that incorporated variation on sample geometry, namely position and orientation, elemental concentration, and density variations formed the training, validation, and test datasets. We investigated four cases that determined the elemental masses for a different number of elements (from a single element up to four elements). The unimodal spectral X-ray radiography and XRF models and the multimodal fusion model were evaluated and compared on each case.

5.2 Materials and Methods

5.2.1 Simulations of Spectral Radiography and XRF Data with Different Targets

Monte Carlo simulations were used to generate the spectral radiography and XRF data. The target position, orientation, and density were randomized across all simulations as described in Subsection 3.2.3. The outputs comprise the following:

- 3D histogram of the spectral radiograph with 80 $\times$ 80 pixels and 160 1-keV bins
- 1D histogram of the XRF spectrum with 160 1-keV bins
- The total mass for each element present in the powder target

Each simulation uses both geantinos and photons to generate the outputs. 10^6 geantinos create the spectral radiograph and determine the elemental masses, and 10^8 photons induce the XRF signal.

An additional simulation was run with only geantinos to generate a Φ_0 spectral radiograph. The cuvette, powder, and XRF detector were removed from this simulation. The resulting data were used for preprocessing of the spectral radiography data.

5.2.2 Unimodal Spectral Radiography Model

The raw spectral radiography simulation data were preprocessed before being used by the deep learning models. First, the spectrum in each pixel was transformed using the method described in Subsubsection 3.1.2 by the HEXITEC response function shown in Figure 3.1. The transformed spectral radiograph was then scaled by a constant to mimic the count rate a 10-min experimental acquisition, and subsequently Poisson noise was added. This yielded a detector response d^{SR} to a 10-min acquisition. The next transformation used d_0^{SR}, the transformed and scaled Φ_0 (with no additional added noise), to create the spectral radiography (SR) input feature X^{SR} given by

$$X^{\text{SR}} = 1 - \frac{d^{\text{SR}}}{d_0^{\text{SR}}} \tag{5.1}$$

The energy dimension was truncated to only include energy bins 40 – 160 keV. The resulting spectral radiography input feature is $X^{\text{SR}} \in \mathbb{R}^{80\times80\times120}$. Figure 5.1 shows a nearly noiseless example that has a W + Bi_2O_3 + Gd_2O_3 powder composition. Along the energy dimension, sharp jumps in intensity are visible that correspond to K-edge energies. There is additional striping around the target edge across the spatial dimensions due to a rotated target and limited spatial resolution.

Figure 5.2 shows four examples of X^{SR} that include Poisson noise, one example for each powder target composition. The noise is most evident at high energies where the flux intensity is small. The spectral features experience blurring effects, and the region around

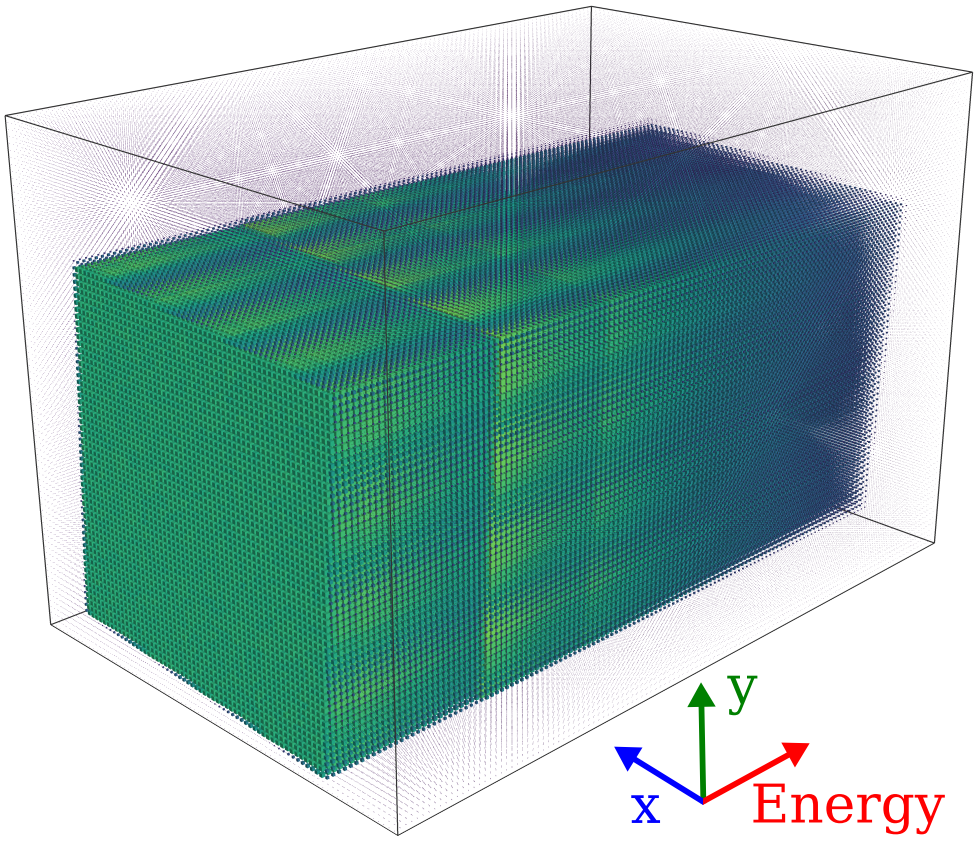

Figure 5.1: The data resulting from spectral X-ray radiography formed 3D tensor characterized by two spatial dimensions and one energy dimension. The spatial dimensions form a radiographic image, and the energy dimension contains characteristic jump discontinuities caused by the energy-dependent attenuation probabilities. These data were generated by simulating the transmission through a W + Bi_2O_3 + Gd_2O_3 target. They were transformed via detector response function and flatfield normalization, and no Poisson noise was added.

the powder visibly gained a baseline noise.

The unimodal architecture was constructed by modifying the ResNet-34 architecture presented in [70]. Designed for image recognition tasks, the ResNet architectures incorporate *skip connections* which provide routes that bypass blocks of convolutional layers. The architecture used is depicted in Figure 5.3 which was adapted to perform 3D convolutions and pooling. This dataset showed increased performance with a deeper network and with more filters, leading to a ResNet-34 structure with 64 filters. An average pooling layer was added to the beginning of the network to reduce the input size. The architecture size was limited by the GPU used for training.

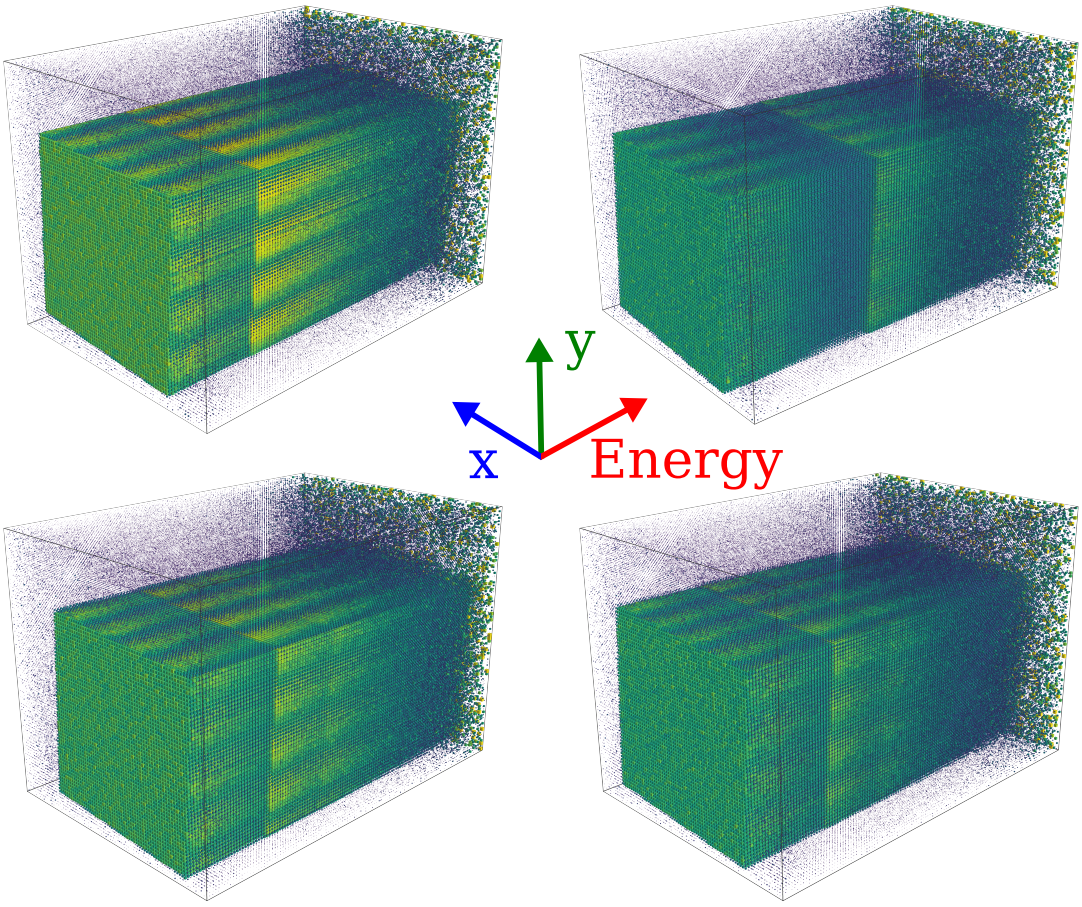

Figure 5.2: These spectral radiographs are have been transformed by the detector response function, by normalizing with flatfield, and by adding Poisson noise. The four examples shown here represent the four types of powder composition used in the simulations: W (top left); Bi_2O_3 (top right); W + Bi_2O_3 (bottom left); W + Bi_2O_3 + Gd_2O_3 (bottom right).

The network was trained to regress the elemental masses by minimizing a mean squared error loss function using an Adam optimizer. We added an additional exponential decay scheduler on the learning rate for fine-tuning the training.

We denote the penultimate layer as $r^{\text{SR}}(X^{\text{SR}}; \theta^{\text{SR}})$, where θ^{SR} are the network parameters. After training, $r^{\text{SR}}(X^{\text{SR}}; \theta^{\text{SR}})$ served as a learned representation or embedding of the input X^{SR}. This representation was extracted as a new feature used during for multimodal fusion.

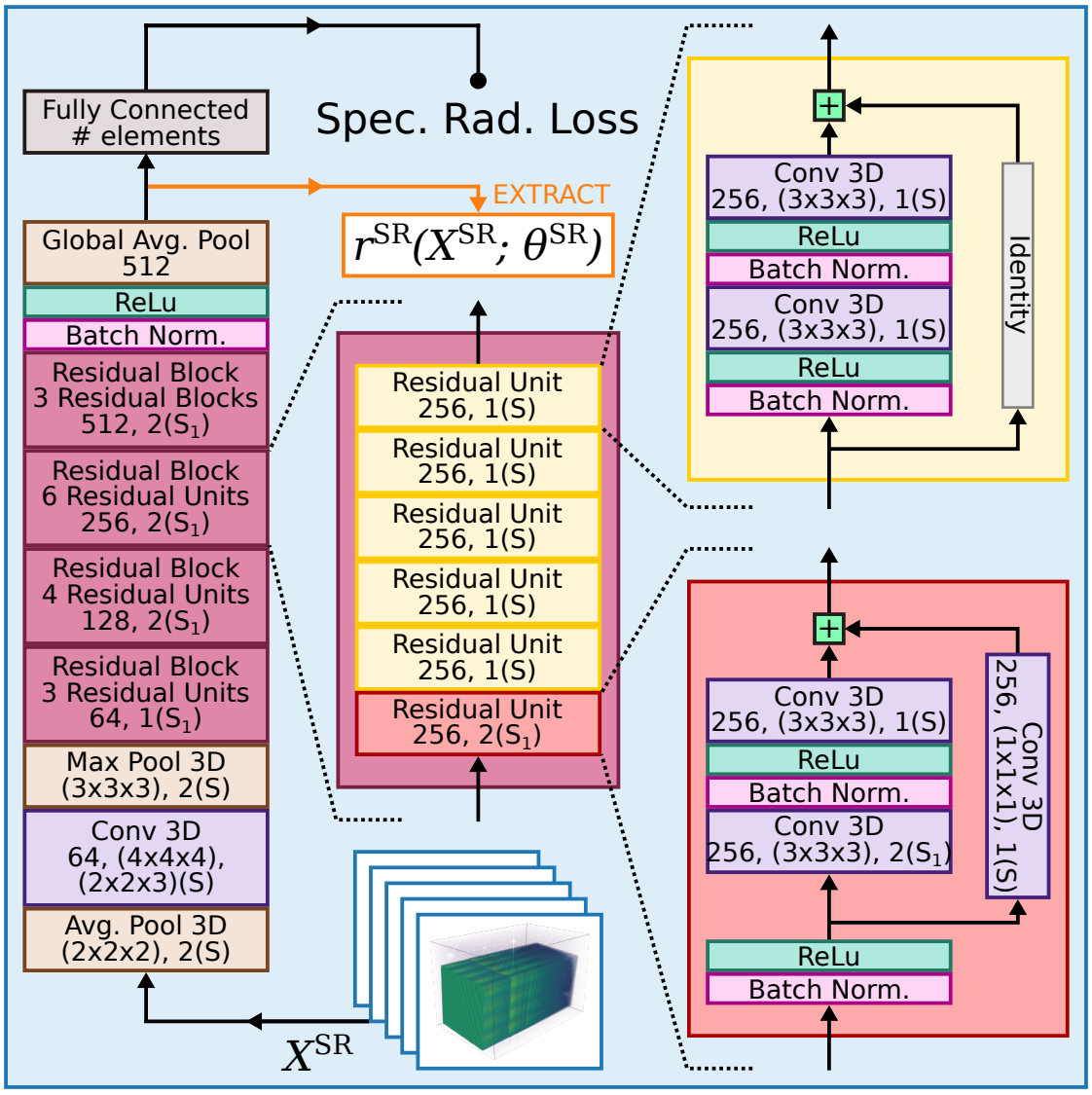

Figure 5.3: The unimodal architecture for spectral X-ray radiography was a 3D ResNet-34. After the initial pooling and stages, the data pass through four residual blocks. A skip connection in a residual block was implemented as summing an identity transformation or a convolutional layer with unitary stride and kernel size. After fully trained, the features $r^{SR}(X^{SR}; \theta^{SR})$ were determined for use by the multimodal model.

5.2.3 Unimodal XRF Model

The raw XRF simulation data were also preprocessed before being used by the deep learning models. Like the spectral radiography data, the XRF spectra were transformed by a detector response function. In this case, we used the Amptek response function shown in

Figure 3.3. We then scaled the transformed XRF spectra and added Poisson noise in the same fashion as the spectral radiography data. Lastly, the XRF spectra were scaled down by a constant as a normalization step. This constant was shared across all simulations. The energy dimension was truncated to only include energy bins 40 – 100 keV. The resulting XRF input feature is $X^{\text{XRF}} \in \mathbb{R}^{600}$, visualized in Figure 5.4.

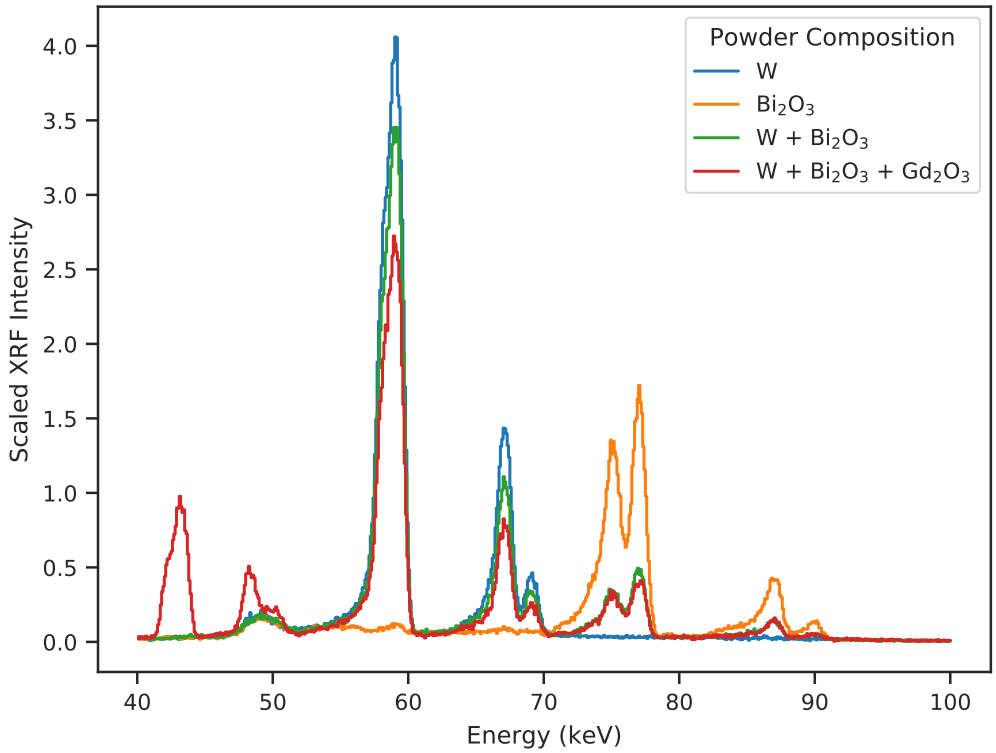

Figure 5.4: The 1D spectrum from the XRF sensor contains characteristic peaks depending on which elements are present and in what quantity. An example from each composition was used to visualize the XRF sensor data.

The unimodal XRF architecture was constructed in a similar way as the unimodal spectral radiography architecture, so we only highlight the differences. This dataset was prone to overfitting, so the chosen network architecture was less complex. Figure 5.5 illustrates the architecture's details. The shallower ResNet-18 structure was used instead with only 4 filters. There was no average pooling added to the start of the XRF network, and average

pooling was used in place of max pooling. The convolutions and pooling were 1D instead of 3D. The training procedure and loss function were the same as the spectral radiography model.

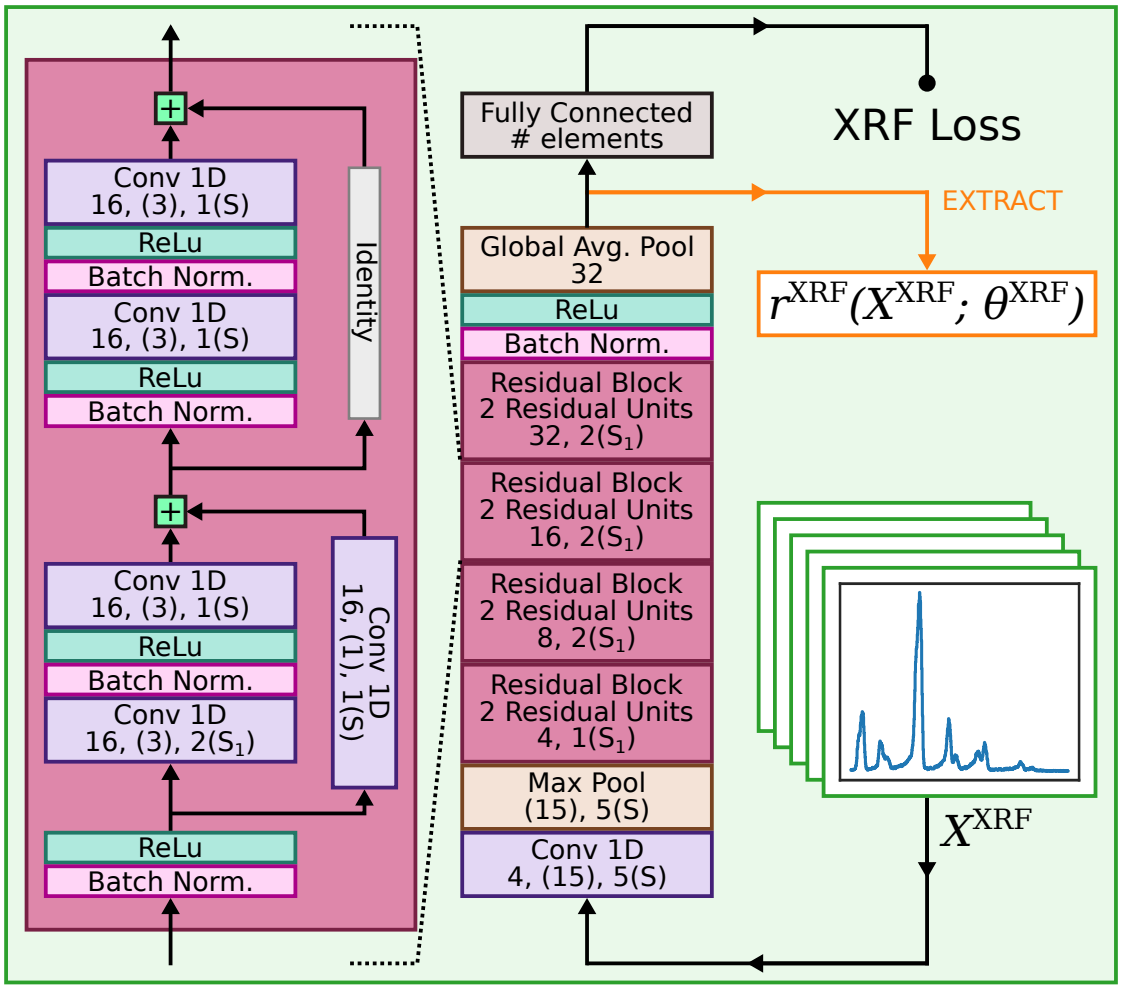

Figure 5.5: The unimodal XRF architecture was also a ResNet structure. The major differences from the spectral X-ray radiography architecture are that the convolutions and pooling were 1D, the residual blocks contained fewer residual units (shallower),and fewer filters were used during in convolutional layers.

Just as in the spectral radiography architecture, the XRF network was used to create a learned representation of the input X^{XRF}. Because the network was smaller, this embedding, $r^{\text{XRF}}(X^{\text{XRF}}; \theta^{\text{XRF}})$, had fewer dimensions than $r^{\text{SR}}(X^{\text{SR}}; \theta^{\text{SR}})$.

5.2.4 Multimodal Fusion Model

The multimodal fusion was performed by first training both unimodal networks individually. The input features for the multimodal (MM) model X^{MM} were formed by concatenating the representations $r^{\text{XRF}}(X^{\text{XRF}}; \theta^{\text{XRF}})$ and $r^{\text{SR}}(X^{\text{SR}}; \theta^{\text{SR}})$, ensuring that each X_i^{XRF} and X_i^{SR} came from the same simulation and therefore have the same labels. The architecture, shown in Figure 5.6, was a densely connected network with four hidden layers. In contrast to the unimodal architectures, the multimodal architecture used a leaky rectified linear (ReLu) activation function and did not use batch normalization. Leaky ReLu is a nonlinear function defined as $f(x) = \max(x, \alpha x)$ where α is some small, positive constant. The loss function, optimizer, and learning rate scheduler were the same as the unimodal models.

5.2.5 Forming Datasets from Sets of Simulations

The four sets of simulations were used to construct the datasets for training the deep learning models. In total, four datasets were created, and a model's task for each dataset differs by the number of elements for which it predicts the mass in the powder. The first dataset (EL1) was made from the set of simulations whose powder target compositions were only W, i.e. one element. The second dataset (EL2) was made from the set of simulations whose powder target compositions were only Bi_2O_3, i.e. two elements. The third dataset (EL3) was made by combining the sets of simulations with the following powder target compositions: W, Bi_2O_3, and W + Bi_2O_3, i.e. three elements. The fourth dataset (EL4) was made by combining the sets of simulations with the following powder target compositions: W, Bi_2O_3, W + Bi_2O_3, and W + Bi_2O_3 + Gd_2O_3, i.e. four elements.

Each set of simulations contained $1.6 \cdot 10^5$ unique simulations with the same powder target composition type. Out of these simulations, 10^5 are designated for training, $2 \cdot 10^4$ were designated for validation, and $2 \cdot 10^4$ were designated for testing. We ran excess simulations as a buffer because a small fraction periodically failed. Consequently, the EL1 and EL2 datasets both had 10^5 training, $2 \cdot 10^4$ validation, and $2 \cdot 10^4$ test examples. The

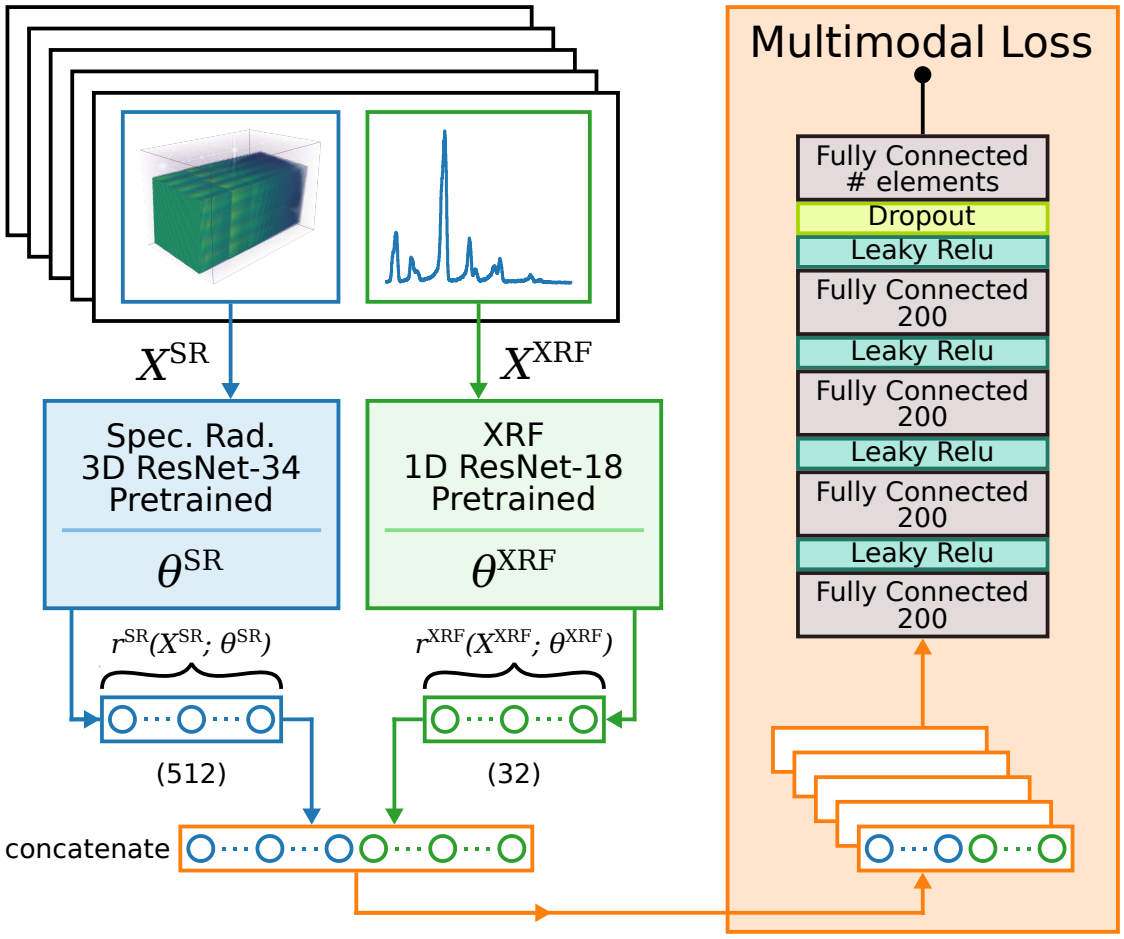

Figure 5.6: The multimodal model first trained the two unimodal models separately. The outputs of the penultimate layers from each unimodal model were evaluated over the dataset, acting as feature extraction. The concatenated features from both models formed the inputs for the multimodal training.

EL3 dataset was a combination of three sets of simulations and consequently had $3 \cdot 10^5$ training, $6 \cdot 10^4$ validation, and $6 \cdot 10^4$ test examples. Similarly, the EL4 dataset had $4 \cdot 10^5$ training, $8 \cdot 10^4$ validation, and $8 \cdot 10^4$ test examples. The sampling schemes for densities and composition are detailed in Appendix C.

5.3 Results

5.3.1 Training and Validation

The training was driven by the minimization of each network's cost function. We used the residuals for each element as performance metrics. Figure 5.7 shows the training and validation residuals of W mass for the EL1 dataset. We monitored the progress of the minimization using the residuals on the training sets for underfitting and convergence. We achieved stable training with a combination of the selected model, its hyperparameters, and the learning rate. The validation residuals provided a measure of the generalization error and overfitting. We concluded that the model was not overfitting the data because the validation did not begin to trend upward.

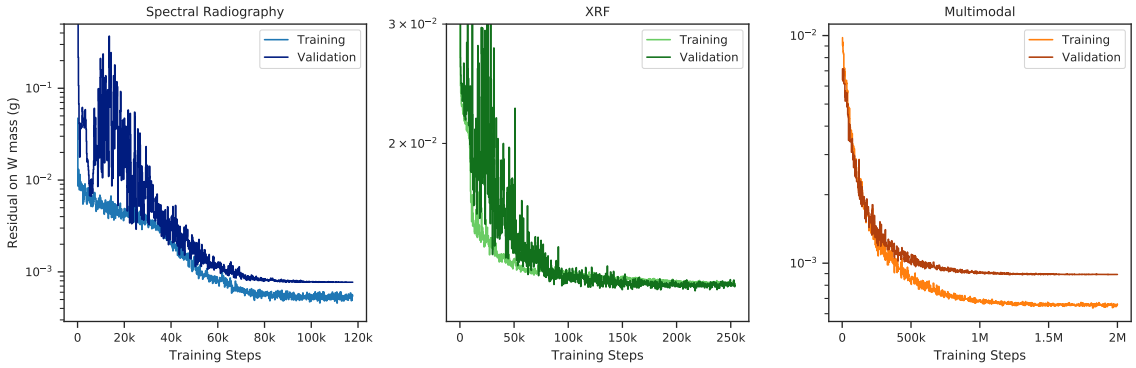

Figure 5.7: W training and validation residuals from EL1 dataset.

The SR and XRF curves behaved differently than the MM network. The two unimodal models started with large validation residuals relative to the training curves and then began approach the training residuals. The MM curves started in close proximity before the validation leveled out and the training continued to decrease. These behaviors were caused by the batch normalization used in the unimodal networks that was absent in the MM network. The *momentum* used in batch normalization delayed the update of the learned mean and variance terms for the validation, while the training set used sample mean and variance of the "current" batch. This process prevented the exploding/vanishing gradients that are

problematic in deep networks like these. Batch normalization additionally provides regularization which helps prevent overfitting. Applying batch normalization to the MM yielded higher training errors and was not used. Instead, dropout was used for regularization.

5.3.2 EL1 Dataset: W

We performed an initial goodness-of-fit test on the three modalities using a two-sample K-S test. The null hypothesis was that the masses determined by the deep learning model come from the same distribution as the mass labels of the outputs. We failed to reject this hypothesis for the SR model (D(1000, 1000) = 0.010, p = 1.000) and for the MM model (D(1000, 1000) = 0.010, p = 1.000). The null hypothesis was rejected for the XRF model (D(1000, 1000) = 0.209, $p < 0.001$).

Residual analysis was used to check the statistical assumptions necessary to use a model for decision making. The scatter plot in Figure 5.8 compares the residuals to the label masses on W data. There was a negative correlation between the XRF residuals and the label masses for W. The residuals from SR and MM models showed no relationship with the mass labels.

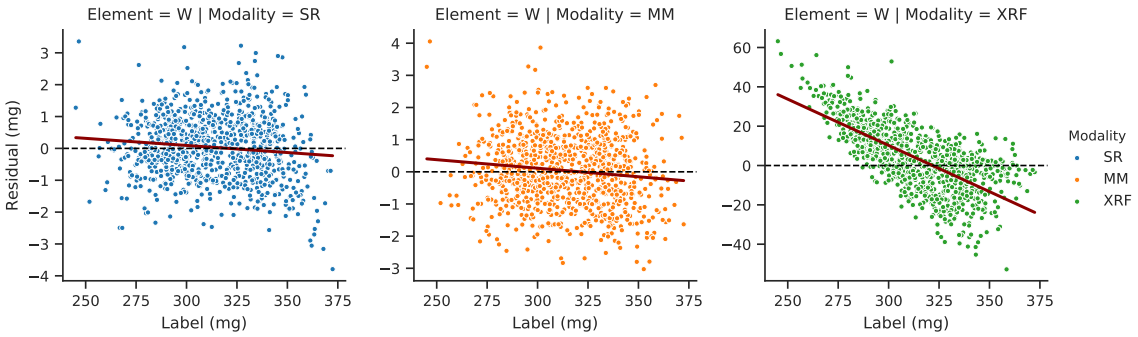

Figure 5.8: Scatter plots of residual vs. label for SR, MM, and XRF models on EL1 dataset for W.

The residual plots for the SR and MM models showed consistent variation across the range of label masses. The residuals from XRF increased in variation with mass, indicating a heteroscedastic distribution. The structure from the XRF residuals cannot be tested or

analyzed in several ways because it breaks these classical assumptions.

We inspected the normality of the SR and MM models with the residual histograms in Figure 5.9. Visual inspection indicated both SR and MM residuals were normal. We tested the null hypothesis that the residuals are normally distributed using a Shapiro-Wilk test of normality. We failed to reject the null hypothesis for all three models (SR: W(1000) = 0.997, p = 0.0457; MM: W(1000) = 0.998, p = 0.354; XRF: W(1000) = 0.998, p = 0.324). Figure 5.10 displays the corresponding probability plots. Both SR and MM showed good agreement with the normality line; all points fell within the 99% confidence bands.

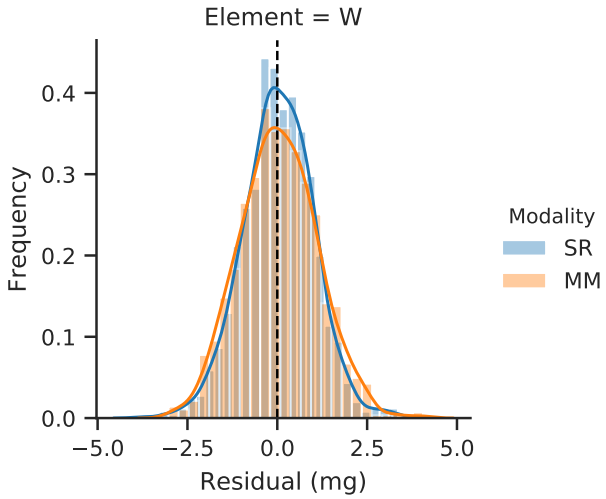

Figure 5.9: Residual histograms for SR and MM models on EL1 dataset.

We limited our investigation of performance to the SR and MM models. The mean of the residuals for SR was 0.0320 mg (99% CI [–0.0473 mg, 0.111 mg]) and for MM was 0.0397 mg (99% CI [–0.0478 mg, 0.127 mg]). There was no statistically significant difference between the residual means from these two modalities as determined by a Welch's t-test (t(1978.898) = –0.168, p = 0.867). The standard deviations of the residuals were 0.971 mg (99% CI [0.917 mg, 1.030 mg]) for SR and 1.071 mg (99% CI [1.012 mg, 1.136 mg]) for MM. There was a statistically significant difference of variances as determined by a Levene test for equal variances (W(1, 1999) = 11.006, $p < 0.001$).

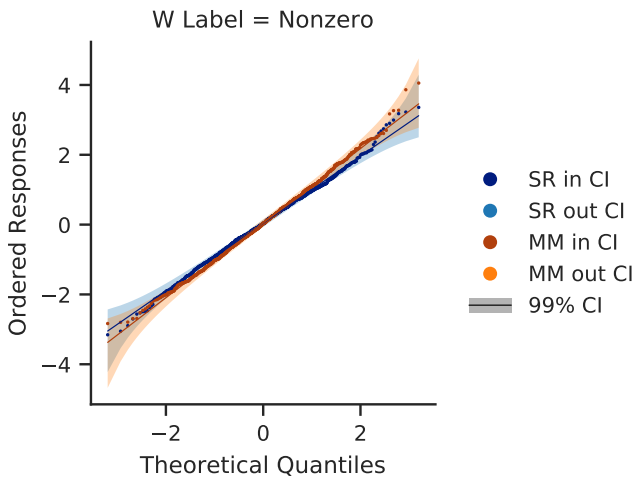

Figure 5.10: Normal probability plots for SR and MM models on EL1 dataset. A strong linear relationship between the ordered responses of the residuals and theoretical quantiles from the standard normal suggests normality.

5.3.3 EL2 Dataset: Bi and O

We performed an initial goodness-of-fit test on the three modalities using a two-sample K-S test. The null hypothesis was that the masses determined by the deep learning model come from the same distribution as the mass labels of the outputs. We failed to reject the null hypothesis for the SR model (D(2000, 2000) = 0.00601, p = 1.000), the MM model (D(2000, 2000) = 0.00501, p = 1.000), and the XRF model (D(2000, 2000) = 0.0531, p = 0.00721).

This dataset was made up of a single compound, Bi_2O_3, containing two elements, Bi and O. The scatter plot in Figure 5.11 compares the residuals to their label masses on Bi and O data. There was a negative correlation between the XRF residuals and the label masses for both Bi and O. The residuals from SR and MM models showed no strong relationship with the mass labels.

The residual plots for the MM model on Bi and O and the spectral radiography model on O showed consistent variation across the range of label masses. The residuals from XRF were constant along the line of correlation but are not constant with the current model. There was a weak, positive dependence of residual variation on label mass for Bi and O on

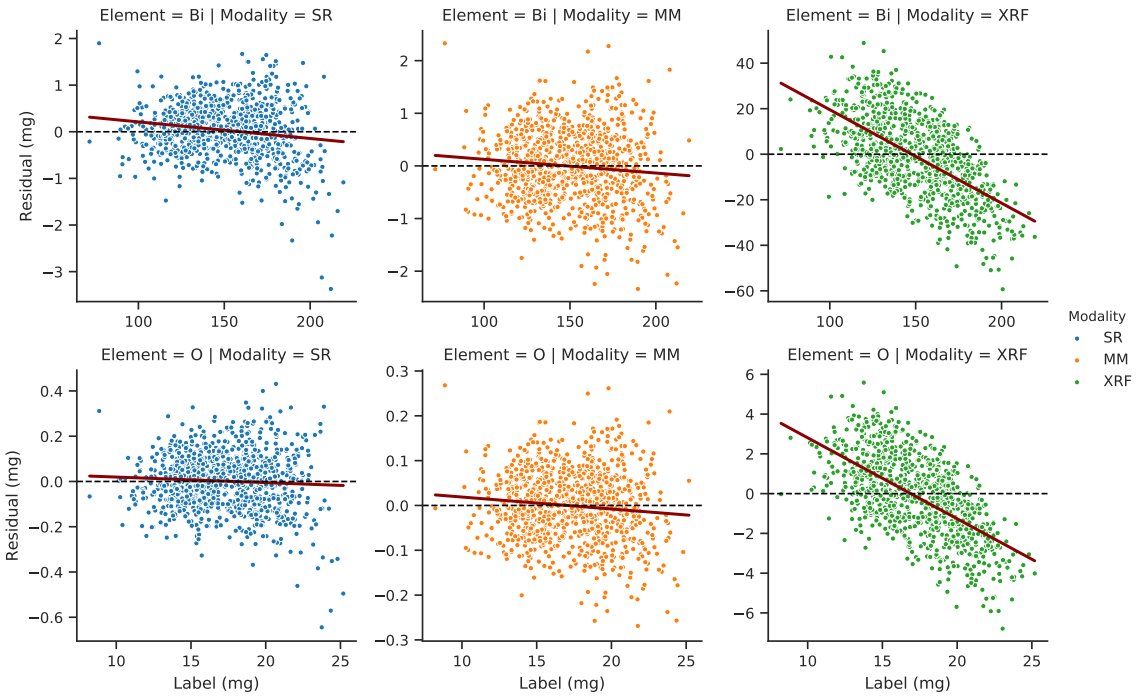

Figure 5.11: Scatter plots of residual vs. label for SR, MM, and XRF models on EL2 dataset for each element (Bi and O).

the SR model.

We inspected the normality of the SR and MM models with the residual histograms in Figure 5.12 which suggests they are not normal. All three modalities were rejected as normal as determined by a Shapiro-Wilk test of normality (SR: W(2000) = 0.902, $p < 0.001$; MM: W(2000) = 0.904, $p < 0.001$; XRF: W(2000) = 0.909, $p < 0.001$). Figure 5.12 exhibited distinct modes that were likely due to the presence of multiple elements.

We then separated the dataset by element to test the null hypothesis that the residuals followed a normal distribution, shown in Figure 5.13. For Bi, we rejected the null hypothesis for the SR model (W(1000) = 0.975, $p < 0.001$), and we failed to reject the null hypothesis for the MM model (W(1000) = 0.998, p = 0.364) and the XRF model (W(1000) = 0.998, p = 0.255). The normal probability plots of the Bi residual in Figure 5.14 show the departure from normality of the SR model with values below the fitted line due to a long tail. For O, we rejected the null hypothesis for the SR model (W(1000) = 0.990, $p < 0.001$) and failed to reject the null hypothesis for the MM model (W(1000) = 0.998, p

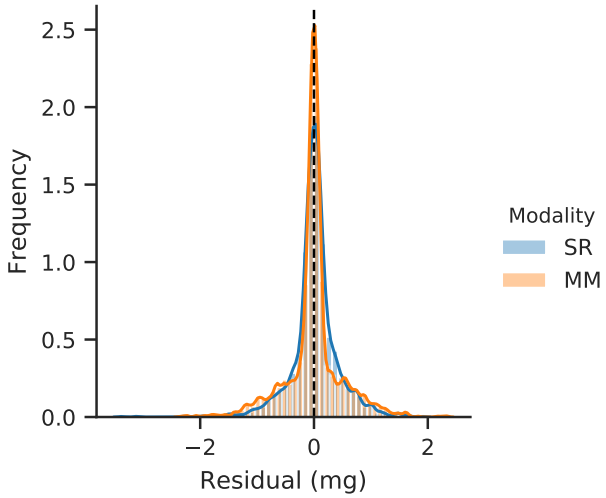

Figure 5.12: Residual histograms for SR and MM models on EL2 dataset.

= 0.356) and the XRF model (W(1000) = 0.998, p = 0.231). The normal probability plots of the O residual in Figure 5.15 show the departure from normality of the SR model with values below the fitted line due to a long tail.

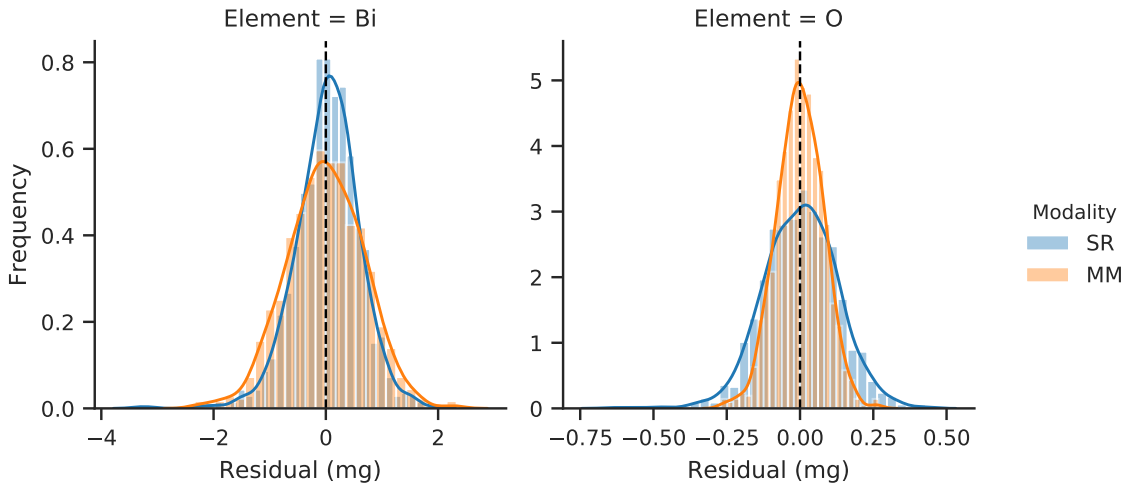

Figure 5.13: Residual histograms for SR and MM models on EL2 dataset, separated by element.

We limited our investigation of performance to the SR and MM models. The unimodal XRF model consistently had significantly higher residual variance.

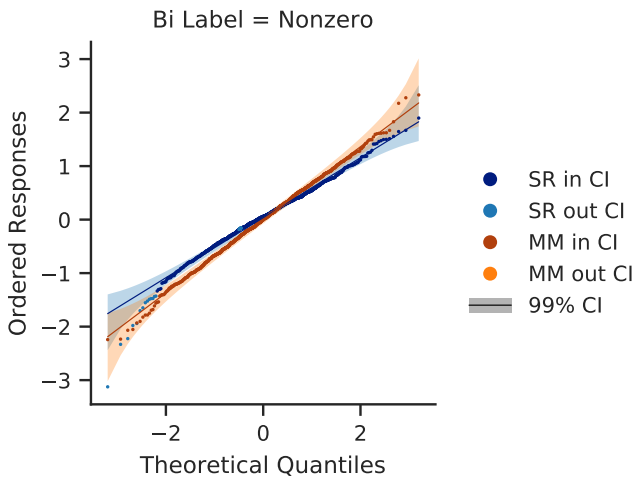

Figure 5.14: Normal probability plots for SR and MM models on Bi in EL2 dataset.

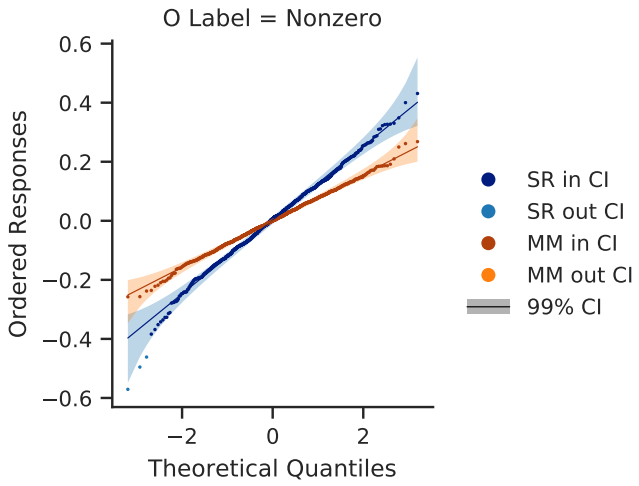

Figure 5.15: Normal probability plots for SR and MM models on O in EL2 dataset.

Performance of SR and MM Models on Bi

The mean of the residuals for SR was 0.0319 mg (99% CI [–0.0150 mg, 0.0788 mg]) and for MM was –0.00628 mg (99% CI [–0.0623 mg, 0.0497 mg]). There was no statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test ($t(1935.974) = 1.350$, $p = 0.177$) and a Mann-Whitney U-test ($U(1000) = 478200$, $p = 0.0534$). The standard deviations of the residuals were 0.574 mg (99% CI [0.543 mg, 0.609 mg]) for SR and 0.686 mg (99% CI [0.648 mg, 0.728 mg]) for MM. There

was statistically significant difference of variances between these modalities as determined by a Levene test for equal variances ($W(1, 1999) = 39.015$, $p < 0.001$).

Performance of SR and MM Models on O

The mean of the residuals for SR was 0.00158 mg (99% CI [– 0.00876 mg, 0.0119 mg]) and for MM was –0.000648 mg (99% CI [–0.00708 mg, 0.00579 mg]). There was no statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test ($t(1670.599) = 0.473$, $p = 0.637$) and a Mann-Whitney U-test ($U(1000) = 489227.5$, $p = 0.224$). The standard deviations of the residuals were 0.127 mg (99% CI [0.120 mg, 0.134 mg]) for SR and 0.0788 mg (99% CI [0.0745 mg, 0.0836 mg]) for MM. There was a statistically significant difference of variances as determined by a Levene test for equal variances ($W(1, 1999) = 156.358$, $p < 0.001$).

5.3.4 EL3 Dataset: W, Bi, and O

We performed an initial goodness-of-fit test on the three modalities using a two-sample K-S test. The null hypothesis was that the masses determined by the deep learning model come from the same distribution as the mass labels of the outputs. This hypothesis was rejected for the SR model ($D(9000, 9000) = 0.306$, $p < 0.001$), the MM model ($D(9000, 9000) = 0.186$, $p < 0.001$), and the XRF model ($D(9000, 9000) = 0.281$, $p < 0.001$). In many cases, the label was zero for one or more element. We performed the same goodness-of-fit test on examples where the label mass was nonzero. After limiting the test to nonzero labels, we failed to reject the null hypothesis for the SR model ($D(6000, 6000) = 0.00283$, $p = 1.000$), the MM model ($D(6000, 6000) = 0.00183$, $p = 1.000$), and the XRF model ($D(6000, 6000) = 0.026$, $p = 0.0346$).

This dataset was made up of multiple targets. These targets were either a single compound W containing only W, a single compound Bi_2O_3 containing Bi and O, or a mixture of two compounds W and Bi_2O_3 containing W, Bi, and O. The scatter plot in Figure 5.16

compares the residuals to their label masses on W, Bi, and O data. The SR residuals were negatively correlated to label mass for all three elements. A large difference in residual variation was also present when mass was zero compared to the nonzero masses. The variation of the W residual from the SR model increased with W mass, but the Bi and O residuals had a more constant variation across nonzero label masses. The XRF residuals did not indicate correlation with label mass for any element, but the variation increased with label mass for all elements. The MM residuals were uncorrelated to label masses, and the variations were less prominent than SR or XRF. For W, the MM residual variations increased with label mass, and for Bi and O, they decreased with label mass.

We inspected the normality of the models using quantitative and graphical techniques. All three modalities were rejected as normal as determined by a Shapiro-Wilk test of normality (SR: W(9000) = 0.946, $p < 0.001$; MM: W(9000) = 0.799, $p < 0.001$; XRF: W(9000) = 0.819, $p < 0.001$). We separated the dataset by element to test for normality, shown in Figure 5.17. For W, all modalities were rejected as normal (SR: W(3000) = 0.880, $p < 0.001$; MM: W(3000) = 0.918, $p < 0.001$; XRF: W(3000) = 0.903, $p < 0.001$). For Bi, all modalities were rejected as normal (SR: W(3000) = 0.968, $p < 0.001$; MM: W(3000) = 0.936, $p < 0.001$; XRF: W(3000) = 0.925, $p < 0.001$). For O, all modalities were rejected as normal (SR: W(3000) = 0.986, $p < 0.001$; MM: W(3000) = 0.880, $p < 0.001$; XRF: W(3000) = 0.918, $p < 0.001$).

Visual inspection of the residual histograms in Figure 5.17 indicated multiple modes within each distribution. The residuals on W mass were further dissected in Figure 5.18 according to powder target composition. The distinct modes were clearly separated after this split; however, we were interested in the performances on the case where the label mass is nonzero and the case where the label mass is zero. This is the split that we made for further investigation.

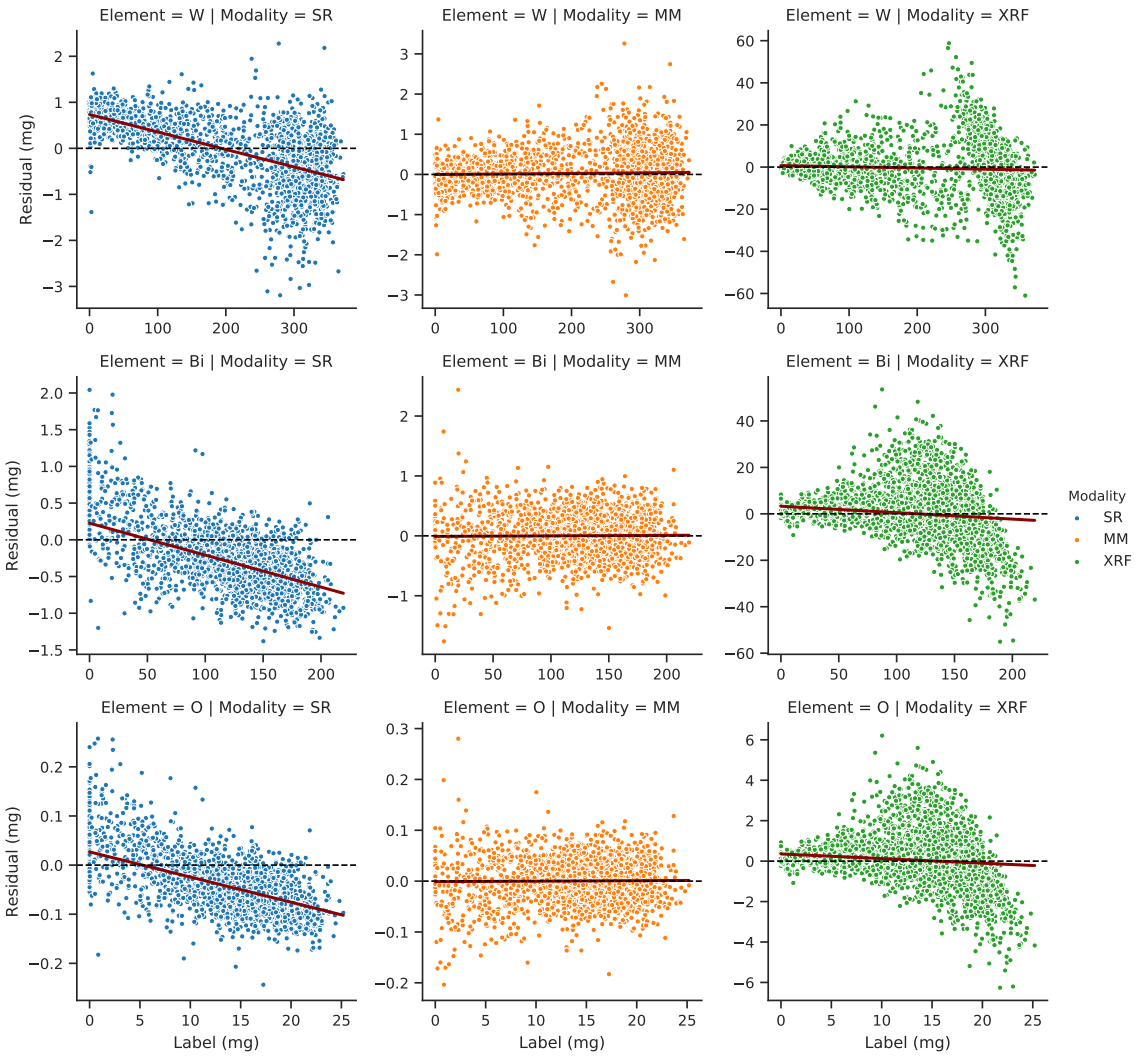

Figure 5.16: Scatter plots of residual vs. label for SR, MM, and XRF models on EL3 dataset for each element (W, Bi, and O).

Performance of SR and MM Models on W

The W residuals are shown in Figure 5.19 and are separated into examples where the W label was zero or nonzero. Using a Shapiro-Wilk test of normality, we rejected the normality hypothesis on all modalities for the data with nonzero W labels (SR: W(2000) = 0.975, $p < 0.001$; MM: W(2000) = 0.987, $p < 0.001$; XRF: W(2000) = 0.986, $p < 0.001$) and for the data with zero W labels (SR: W(1000) = 0.986, $p < 0.001$; MM: W(1000) = 0.741, $p < 0.001$; XRF: W(1000) = 0.913, $p < 0.001$). The normal probability plots of the W residual

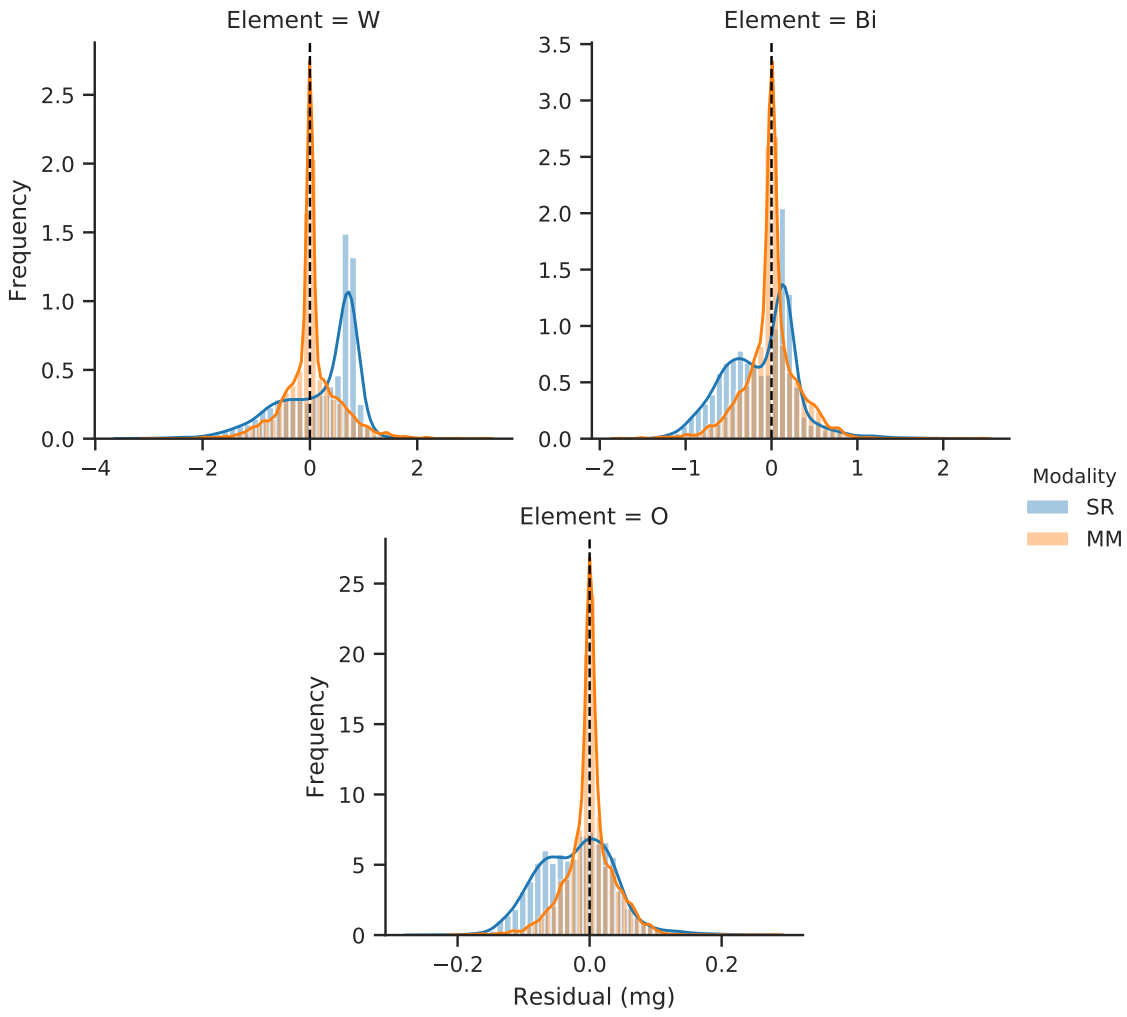

Figure 5.17: Residual histograms for SR and MM models on EL3 dataset.

in Figure 5.20 for nonzero W labels indicated large departures from normality. For the zero W labels, the normal probability plots indicated long tails.

Nonzero Labels. The mean of the residuals for SR was –0.0905 mg (99% CI [–0.133 mg, –0.0476 mg]) and for MM was 0.0236 mg (99% CI [–0.0120 mg, 0.0593 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test ($t(3866.959) = -5.274$, $p < 0.001$), but the difference was not statistically significant as determined by a Mann-Whitney U-test ($U(2000) = 1898116.5$, $p = 0.00264$) which does not assume normality like the t-test. The standard deviations of the residuals were 0.744 mg (99% CI [0.715 mg, 0.776 mg]) for SR and 0.618

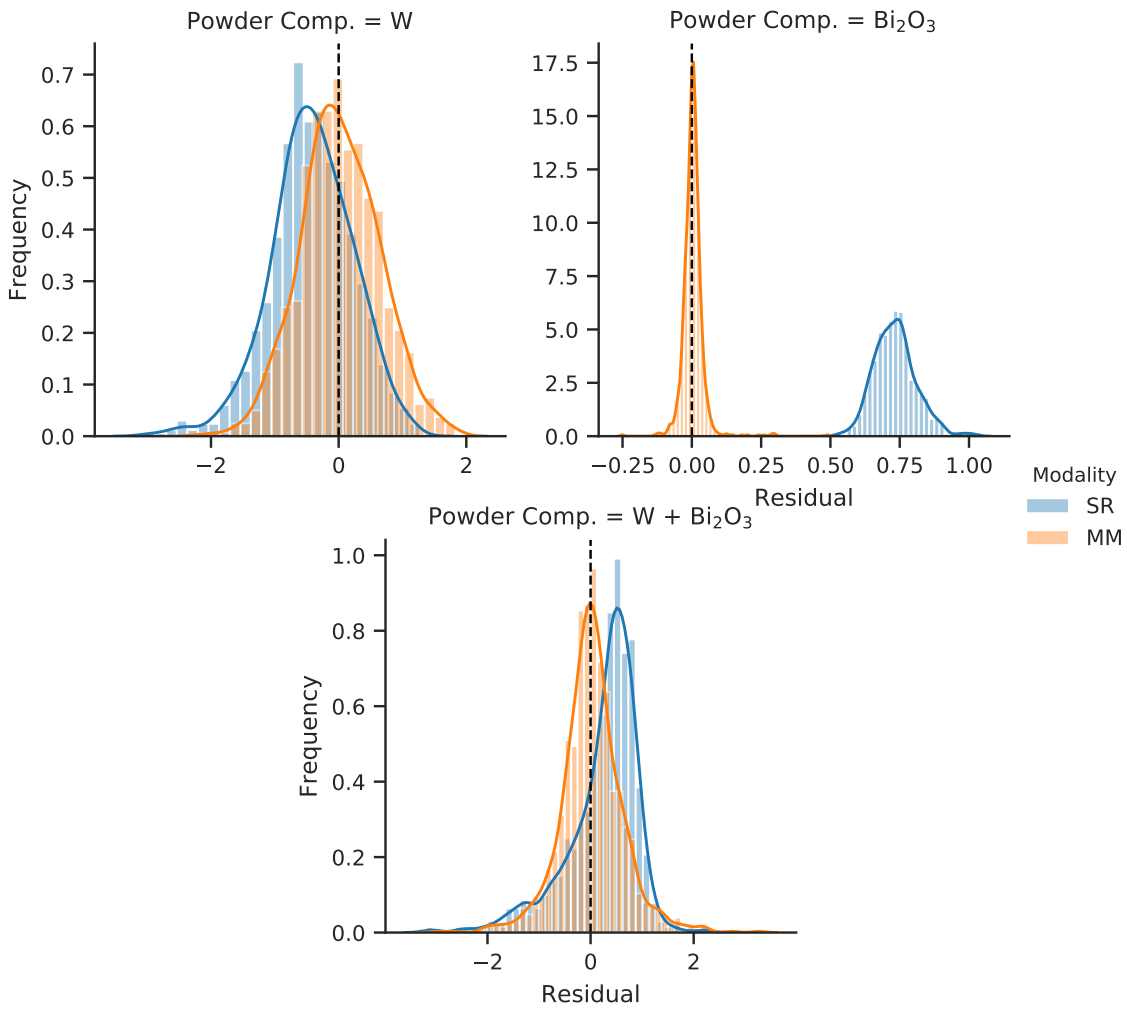

Figure 5.18: W residual histograms for SR and MM models on EL3 dataset separated into the three powder target compositions of EL3: W, Bi_2O_3, and W + Bi_2O_3.

mg (99% CI [0.594 mg, 0.644 mg]) for MM. There was a statistically significant difference of variances as determined by a Levene test for equal variances ($W(1, 3999) = 105.724$, $p < 0.001$).

Zero Labels. The mean of the residuals for SR was 0.733 mg (99% CI [0.727 mg, 0.739 mg]) and for MM was 0.00395 mg (99% CI [0.000726 mg, 0.00717 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test ($t(1506.409) = 270.346$, $p < 0.001$) and a Mann-Whitney U-test ($U(1000) = 0$, $p < 0.001$). The standard deviations of the residuals were 0.0755 mg

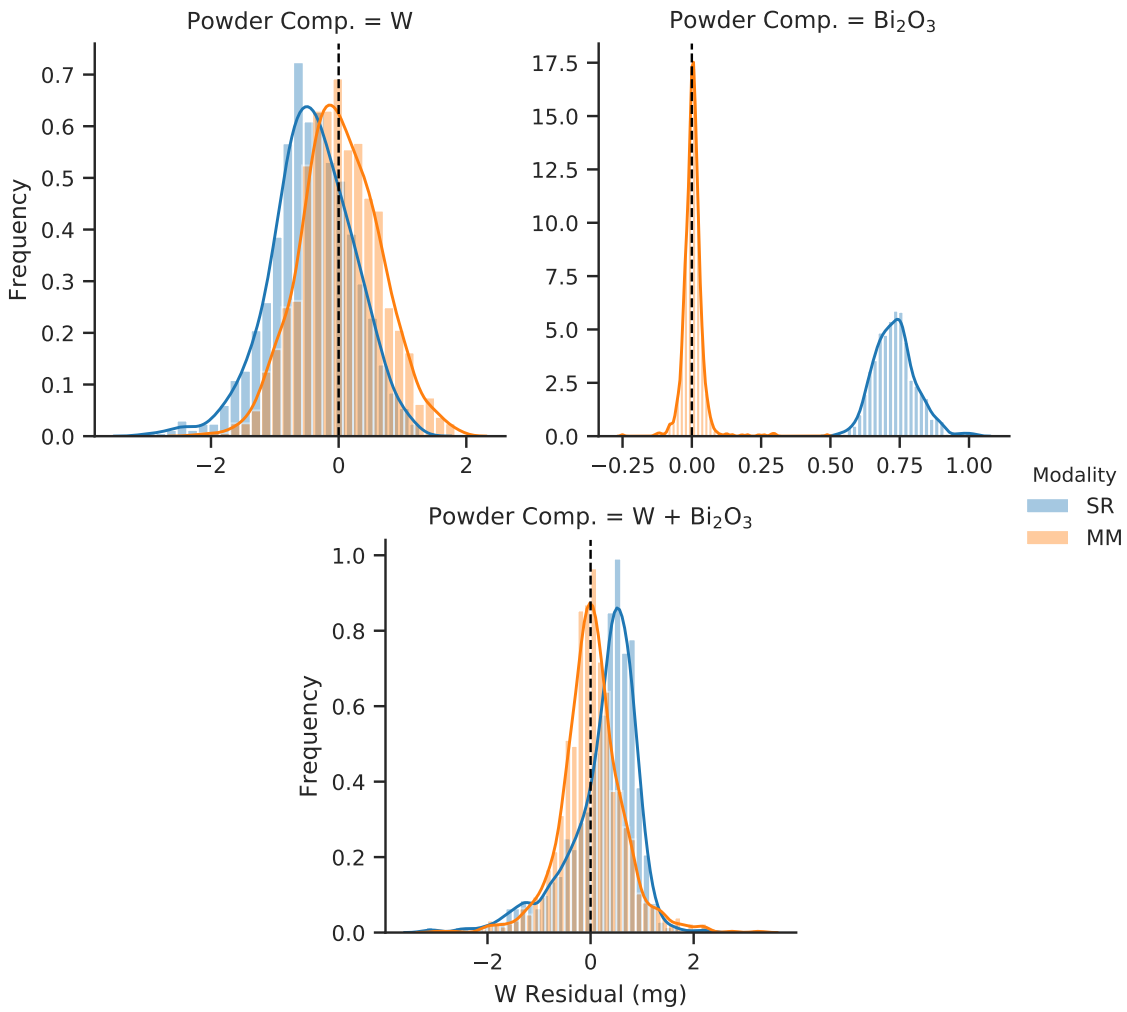

Figure 5.19: W residual histograms for SR and MM models on EL3 dataset separated by zero and nonzero W labels.

(99% CI [0.0714 mg, 0.0801 mg]) for SR and 0.0395 mg (99% CI [0.0373 mg, 0.0419 mg]) for MM. There was a statistically significant difference of variances as determined by a Levene test for equal variances ($W(1, 1999) = 391.457$, $p < 0.001$).

Performance of SR and MM Models on Bi

The Bi residuals are shown in Figure 5.21 and are separated into examples where the Bi label was zero or nonzero. Using a Shapiro-Wilk test of normality, we rejected the normality hypothesis for the nonzero Bi labels (SR: $W(2000) = 0.966$, $p < 0.001$; MM: W(2000)

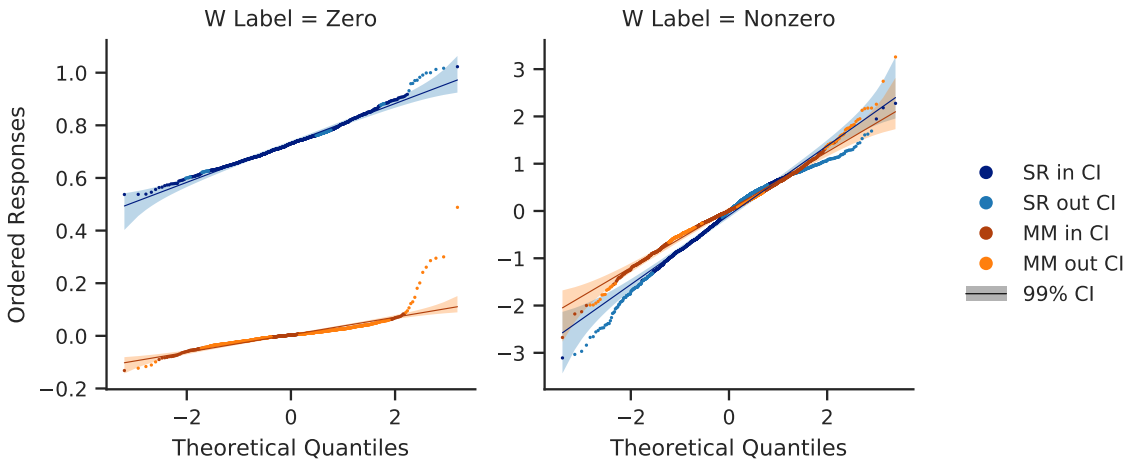

Figure 5.20: Probability plots for SR and MM models on W in EL3 dataset.

= 0.987, $p < 0.001$; XRF: W(2000) = 0.994, $p < 0.001$) and for the zero Bi labels (SR: W(1000) = 0.571, $p < 0.001$; MM: W(1000) = 0.680, $p < 0.001$; XRF: W(1000) = 0.791, $p < 0.001$). The normal probability plots of the Bi residual in Figure 5.22 for nonzero Bi labels indicated long tails where the SR showed more departure than MM. For the zero Bi labels, the normal probability plot for SR showed a strong, nonlinear relationship indicating a skewed distribution. The MM plot showed signs of long tails.

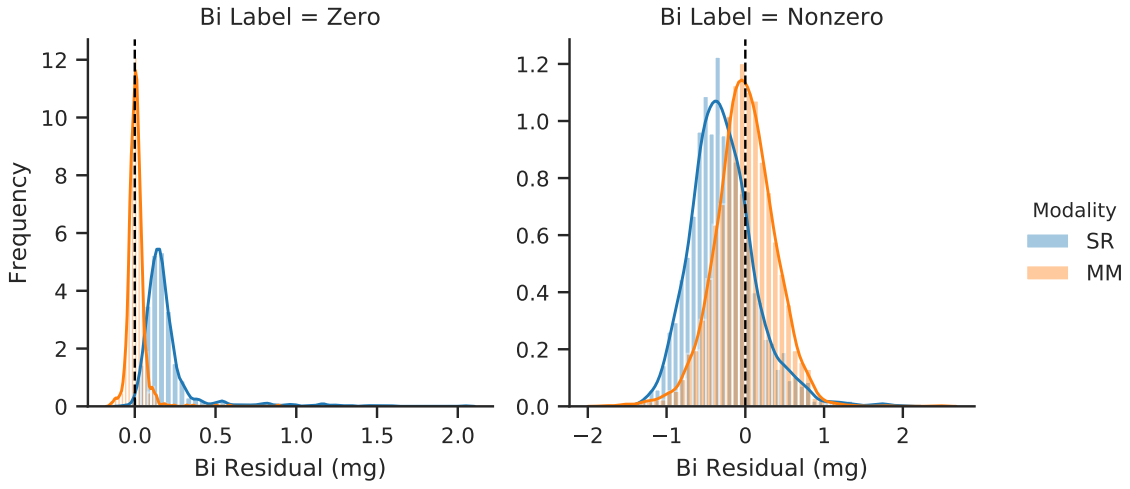

Figure 5.21: Bi residual histograms for SR and MM models on EL3 dataset separated by zero and nonzero Bi labels.

Nonzero Labels. The mean of the residuals for SR was –0.284 mg (99% CI [–0.308

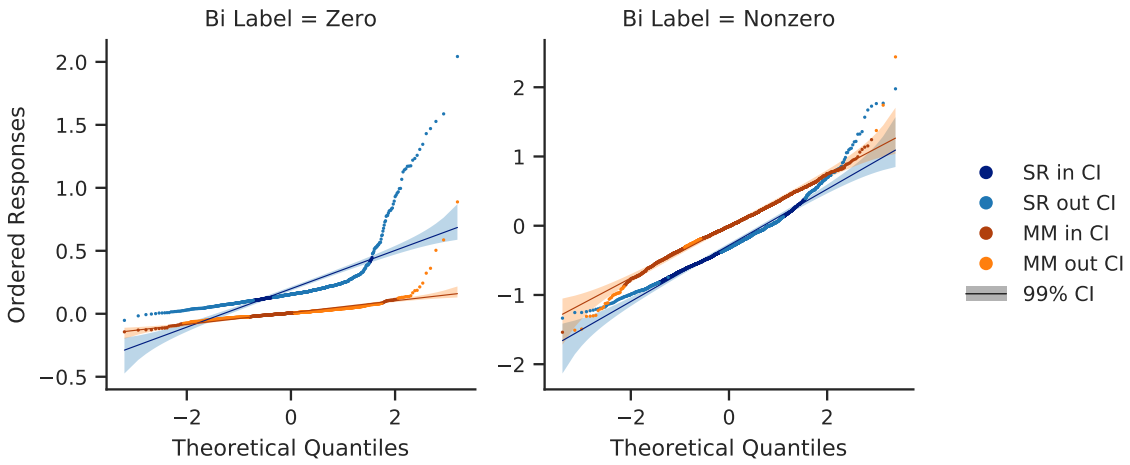

Figure 5.22: Probability plots for SR and MM models on EL3 dataset.

mg, –0.260 mg]) and for MM was –0.00746 mg (99% CI [–0.0293 mg, 0.0144 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test ($t(3968.301) = -22.045$, $p < 0.001$) and a Mann-Whitney U-test ($U(2000) = 1141725$, $p < 0.001$). The standard deviations of the residuals were 0.413 mg (99% CI [0.397 mg, 0.431 mg]) for SR and 0.379 mg (99% CI [0.364 mg, 0.395 mg]) for MM. There was no statistically significant difference of variances as determined by a Levene test for equal variances ($W(1, 3999) = 7.543$, $p = 0.00605$).

Zero Labels. The mean of the residuals for SR was 0.198 mg (99% CI [0.182 mg, 0.215 mg]) and for MM was 0.00901 mg (99% CI [0.00432 mg, 0.0137 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test ($t(1159.557) = 287.448$, $p < 0.001$) and a Mann-Whitney U-test ($U(1000) = 32288$, $p < 0.001$). The standard deviations of the residuals were 0.202 mg (99% CI [0.191 mg, 0.214 mg]) for SR and 0.0574 mg (99% CI [0.0543 mg, 0.0609 mg]) for MM. There was statistically significant difference of variances as determined by a Levene test for equal variances ($W(1, 1999) = 114.666$, $p < 0.001$).

Performance of SR and MM Models on O

The O residuals are shown in Figure 5.23 and are separated into examples where the O label was zero or nonzero. Using a Shapiro-Wilk test of normality, we rejected the normality hypothesis for the nonzero O labels (SR: W(2000) = 0.953, $p < 0.001$; MM: W(2000) = 0.987, $p < 0.001$; XRF: W(2000) = 0.994, $p < 0.001$) and for the zero O labels (SR: W(1000) = 0.906, $p < 0.001$; MM: W(1000) = 0.745, $p < 0.001$; XRF: W(1000) = 0.795, $p < 0.001$). The normal probability plots in Figure 5.24 from SR were nonlinear and indicated skewed distributions. The deviations from MM were indicative of long tails.

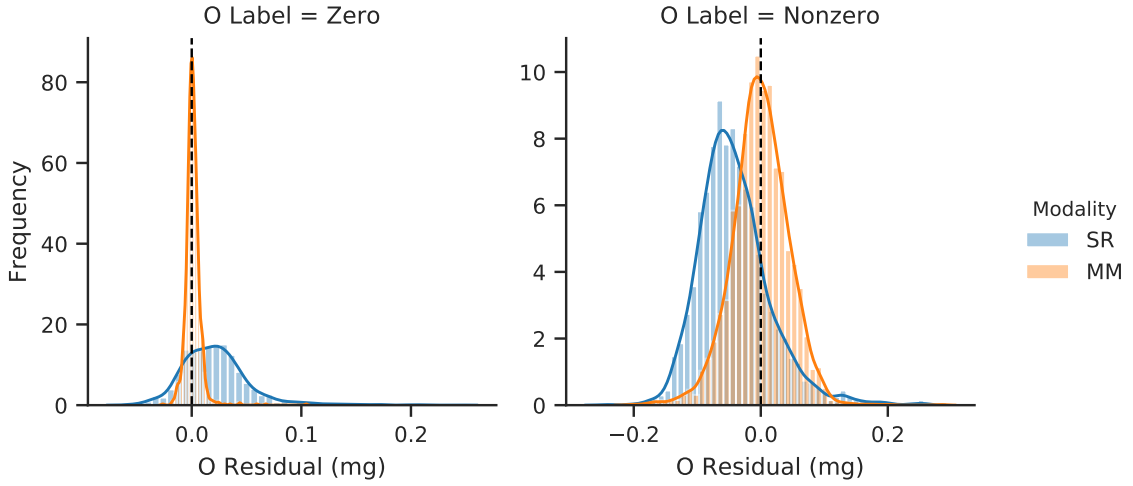

Figure 5.23: O residual histograms for SR and MM models on EL3 dataset separated by zero and nonzero O labels.

Nonzero Labels. The mean of the residuals for SR was –0.0414 mg (99% CI [–0.0447 mg, –0.0382 mg]) and for MM was –0.000865 mg (99% CI [–0.00341 mg, 0.00168 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test ($t(3788.490) = -25.380$, $p < 0.001$) and a Mann-Whitney U-test ($U(2000) = 1013861$, $p < 0.001$). The standard deviations of the residuals were 0.0562 mg (99% CI [0.0540 mg, 0.0585 mg]) for SR and 0.0442 mg (99% CI [0.0425 mg, 0.0461 mg]) for MM. There was statistically significant difference of variances as determined by a Levene test for equal variances ($W(1, 3999) = 57.970$, $p < 0.001$).

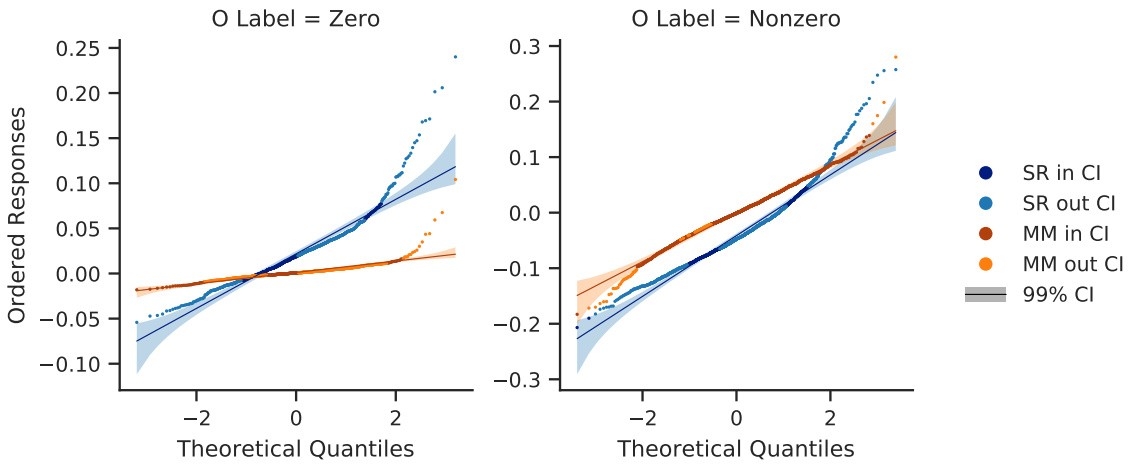

Figure 5.24: Probability plots for SR and MM models on O in EL3 dataset.

Zero Labels. The mean of the residuals for SR was 0.0217 mg (99% CI [0.0191 mg, 0.0243 mg]) and for MM was 0.00113 mg (99% CI 0.000525 mg, 0.00173 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test ($t(1106.708) = 19.880$, $p < 0.001$) and a Mann-Whitney U-test ($U(1000) = 252454$, $p < 0.001$). The standard deviations of the residuals were 0.0318 mg (99% CI [0.0301 mg, 0.0337 mg]) for SR and 0.00740 mg (99% CI [0.00699 mg, 0.00785 mg]) for MM. There was statistically significant difference of variances as determined by a Levene test for equal variances ($W(1, 1999) = 607.345$, $p < 0.001$).

5.3.5 EL4 Dataset: W, Bi, O, and Gd

We performed an initial goodness-of-fit test on the three modalities using a two-sample K-S test. The null hypothesis was that the masses determined by the deep learning model come from the same distribution as the mass labels of the outputs. This hypothesis was rejected for the SR model ($D(16000, 16000) = 0.356$, $p < 0.001$), for the MM model ($D(16000, 16000) = 0.194$, $p < 0.001$), and for the XRF model ($D(16000, 16000) = 0.188$, $p < 0.001$). In many cases, the label is zero for one or more element. We performed the same goodness-of-fit test on examples where the label mass was nonzero. After limiting the test to nonzero labels, the hypothesis was accepted for the SR model ($D(10000, 10000)$

= 0.00420, p = 1.000), for the MM model (D(10000, 10000) = 0.00220, p = 1.000), and for the XRF model (D(10000, 10000) = 0.0238, p = 0.00693).

This dataset was made up of multiple targets. These targets were one of the following: a single compound containing only W, a single compound Bi_2O_3 containing Bi and O, a mixture of two compounds W and Bi_2O_3 containing W, Bi, and O, or a mixture of three compounds W, Bi_2O_3, and Gd_2O_3 containing W, Bi, O, and Gd. The scatter plot in Figure 5.25 compares the residuals to their label masses on W, Bi, O, and Gd data. The SR residuals were negatively correlated to label mass for W, Bi, and O but no correlation with Gd label mass. A large difference in residual variation was also present when label mass was zero compared to the nonzero masses. The variation of the W residual from the SR model increased with W mass, and the Bi and O residuals had roughly constant variation across nonzero label masses. The XRF residuals did not indicate strong correlation with label mass for any element. The variation increased with label mass for W, Bi, and O and was roughly constant for Gd. The MM residuals were uncorrelated to label masses, and the variations were more constant over label masses than SR or XRF. For W, the MM residual variations increased with label mass; for Bi and O, they decreased with label mass; for Gd, they were roughly constant.

We inspected the normality of the models using quantitative and graphical techniques. All three were rejected as normal as determined by a Shapiro-Wilk test of normality (SR: W(16000) = 0.875, $p < 0.001$; MM: W(16000) = 0.813, $p < 0.001$; XRF: W(16000) = 0.717, $p < 0.001$). We separated the dataset by element to test for normality, shown in Figure 5.26. All modalities were rejected as normal for W (SR: W(4000) = 0.878, $p < 0.001$; MM: W(4000) = 0.944, $p < 0.001$; XRF: W(4000) = 0.900, $p < 0.001$), for Bi (SR: W(4000) = 0.961, $p < 0.001$; MM: W(4000) = 0.969, $p < 0.001$; XRF: W(4000) = 0.907, $p < 0.001$), for O (SR: W(4000) = 0.955, $p < 0.001$; MM: W(4000) = 0.959, $p < 0.001$; XRF: W(4000) = 0.919, $p < 0.001$), and for Gd (SR: W(4000) = 0.910, $p < 0.001$; MM: W(4000) = 0.707, $p < 0.001$; XRF: W(4000) = 0.671, $p < 0.001$).

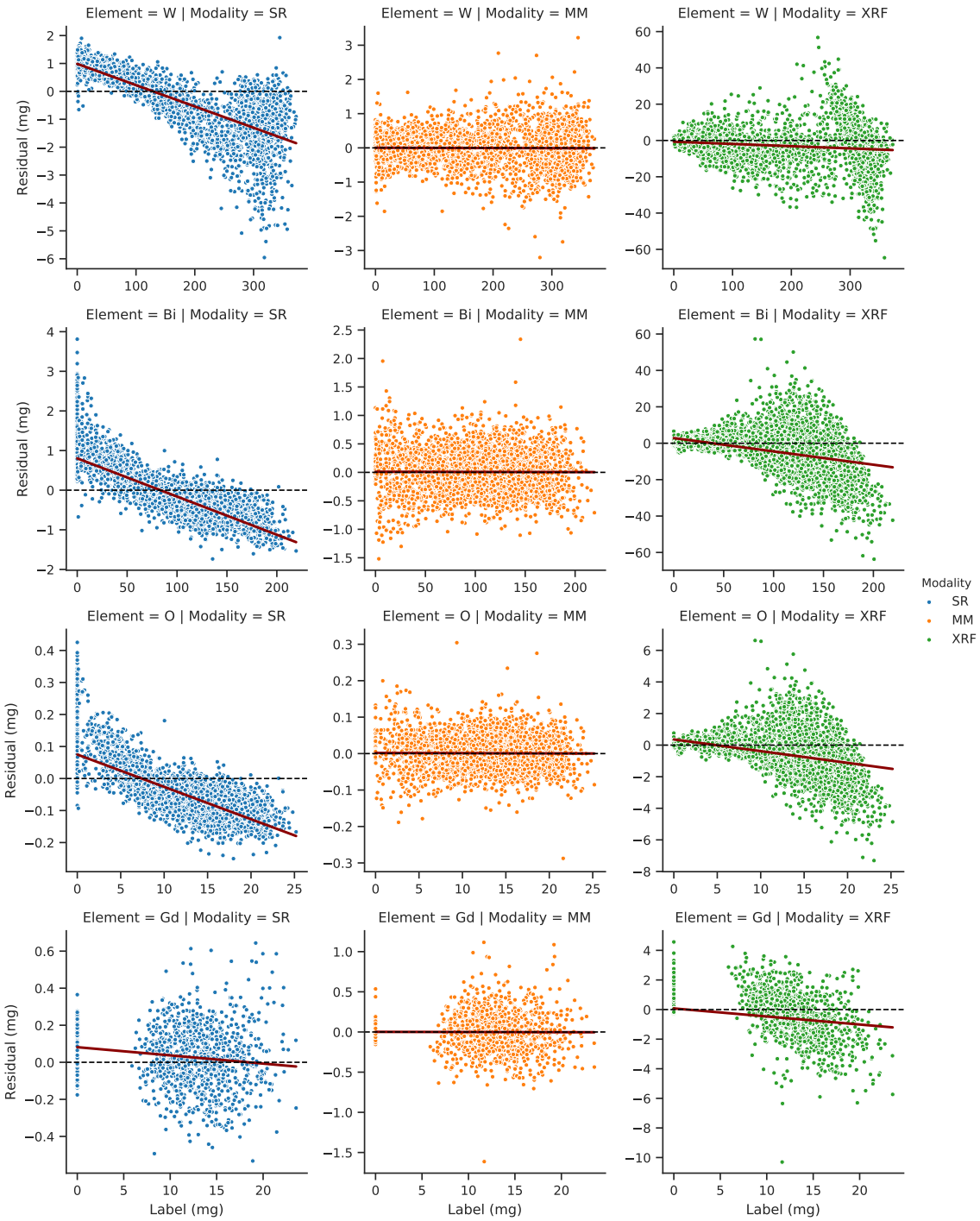

Figure 5.25: Scatter plots of residual vs. label for SR, MM, and XRF models on EL4 dataset for each element (W, Bi, O, and Gd).

Visual inspection of the residual histograms in Figure 5.26 indicated multiple modes within each distribution. The residuals on Bi mass were further dissected in Figure 5.27

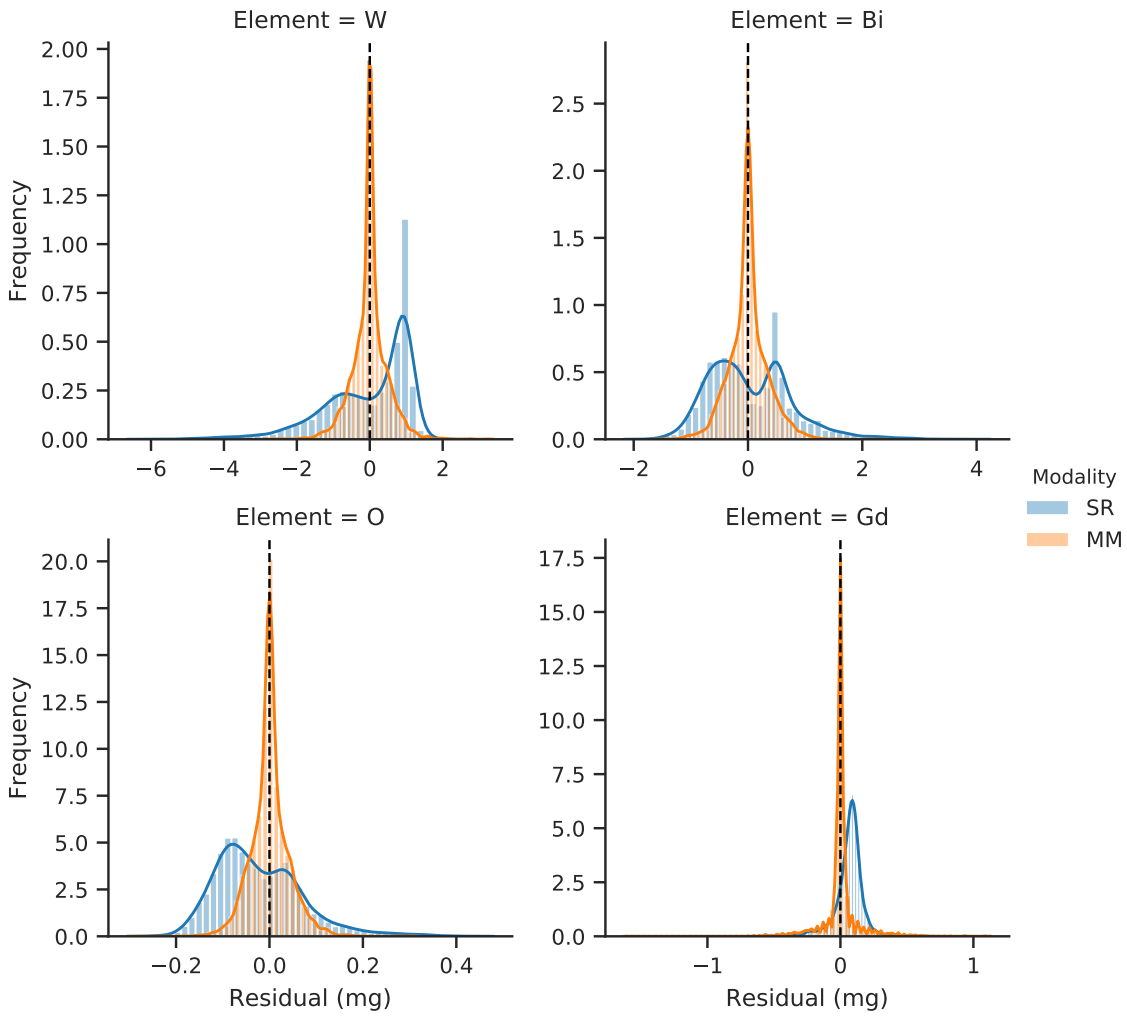

Figure 5.26: Residual histograms for SR and MM models on EL4 dataset.

according to powder target composition. The distinct modes were clearly separated after this split. We split each element's residual distribution by whether the label mass was zero or nonzero as we did for EL3.

Performance of SR and MM Models on W

The W residuals are shown in Figure 5.28 and are separated into examples where the W label was zero or nonzero. Using a Shapiro-Wilk test of normality, we rejected the normality hypothesis for the nonzero W labels (SR: W(3000) = 0.912, $p < 0.001$; MM: W(3000) = 0.983, $p < 0.001$; XRF: W(3000) = 0.961, $p < 0.001$) and for the zero W labels (SR:

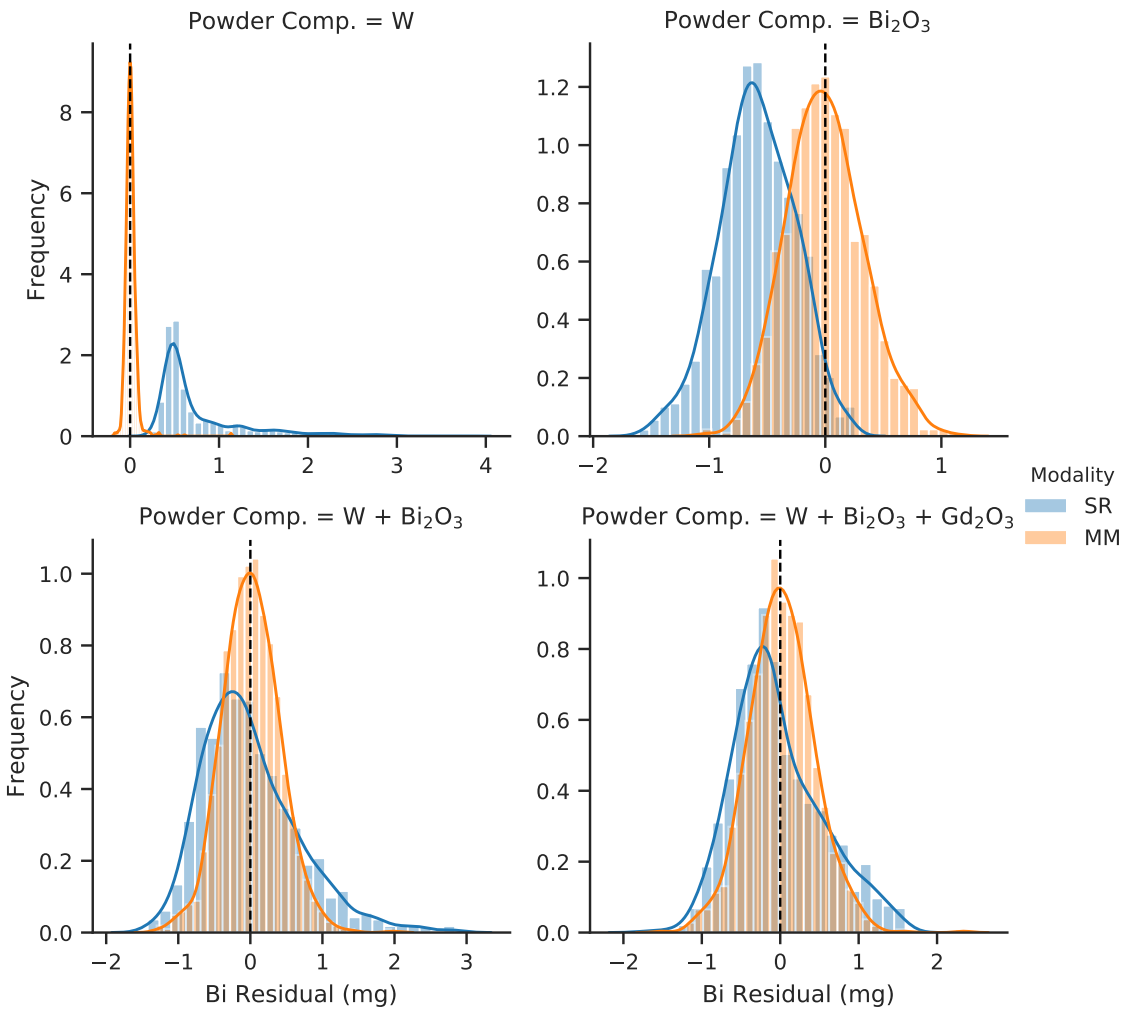

Figure 5.27: Bi residual histograms for SR and MM models on EL4 dataset separated into the four powder target compositions of EL4: W, Bi_2O_3, and W + Bi_2O_3, W + Bi_2O_3 + Gd_2O_3.

W(1000) = 0.972, $p < 0.001$; MM: W(1000) = 0.583, $p < 0.001$; XRF: W(1000) = 0.944, $p < 0.001$). The normal probability plots in Figure 5.29 show that SR had small departures from normality on zero W labels but was strongly skewed on nonzero W labels. The MM model showed signs of long-tails departures.

Nonzero Labels. The mean of the residuals for SR was –0.382 mg (99% CI [–0.437 mg, –0.327 mg]) and for MM was –0.00505 mg (99% CI [–0.0307 mg, 0.0206 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test (t(4242.721) = –16.006, $p < 0.001$) and a Mann-Whitney

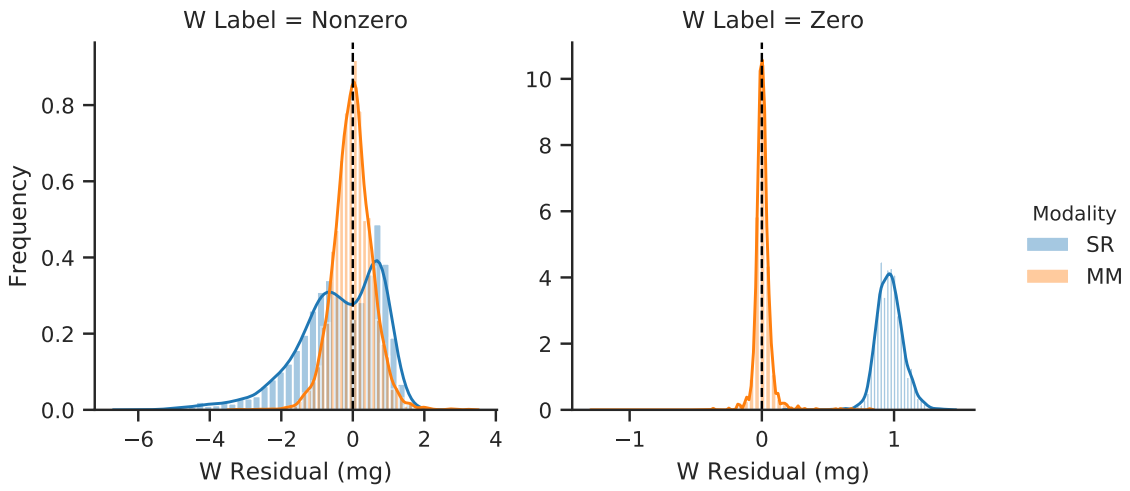

Figure 5.28: W residual histograms for SR and MM models on EL4 dataset separated by zero and nonzero W labels.

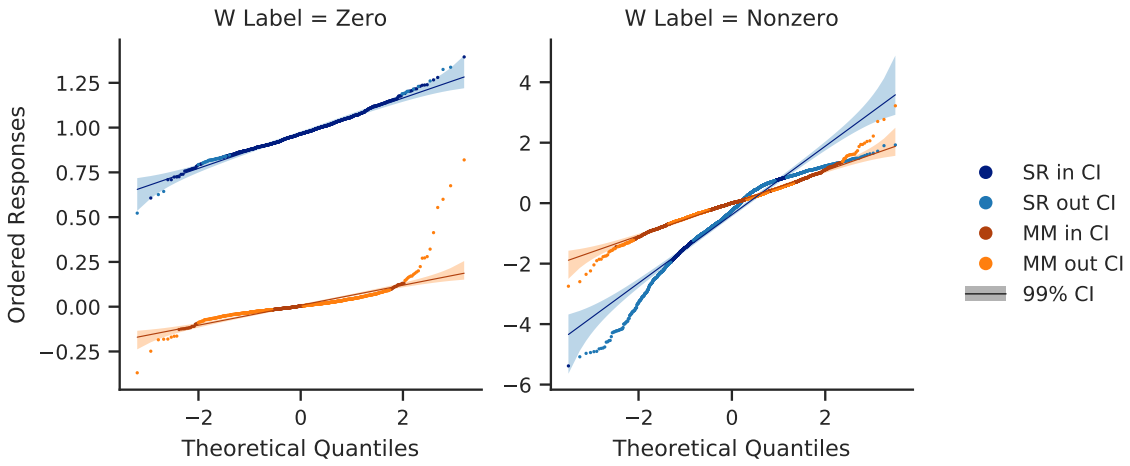

Figure 5.29: Probability plots for SR and MM models on W in EL4 dataset.

U-test (U(3000) = 3862167.5, $p < 0.001$). The standard deviations of the residuals were 1.169 mg (99% CI [1.131 mg, 1.209 mg]) for SR and 0.656 mg (99% CI [0.527 mg, 0.563 mg]) for MM. There was statistically significant difference of variances as determined by a Levene test for equal variances (W(1, 5999) = 1250.995, $p < 0.001$).

Zero Labels. The mean of the residuals for SR was 0.968 mg (99% CI [0.960 mg, 0.976 mg]) and for MM was 0.00706 mg (99% CI [0.000652 mg, 0.0135 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test (t(1880.872) = 237.149, $p < 0.001$) and a Mann-Whitney

U-test (U(1000) = 56, $p < 0.001$). The standard deviations of the residuals were 0.101 mg (99% CI [0.0957 mg, 0.107 mg]) for SR and 0.0785 mg (99% CI [0.0742 mg, 0.0832 mg]) for MM. There was statistically significant difference of variances as determined by a Levene test for equal variances (W(1, 1999) = 163.527, $p < 0.001$).

Performance of SR and MM Models on Bi

The Bi residuals are shown in Figure 5.30 and are separated into examples where the Bi label was zero or nonzero. Using a Shapiro-Wilk test of normality, we rejected the normality hypothesis for the nonzero Bi labels (SR: W(3000) = 0.946, $p < 0.001$; MM: W(3000) = 0.994, $p < 0.001$; XRF: W(3000) = 0.970, $p < 0.001$) and for the zero Bi labels (SR: W(1000) = 0.718, $p < 0.001$; MM: W(1000) = 0.591, $p < 0.001$; XRF: W(1000) = 0.937, $p < 0.001$). The normal probability plots in Figure 5.31 show SR had large departures from normality and MM had long-tail departures.

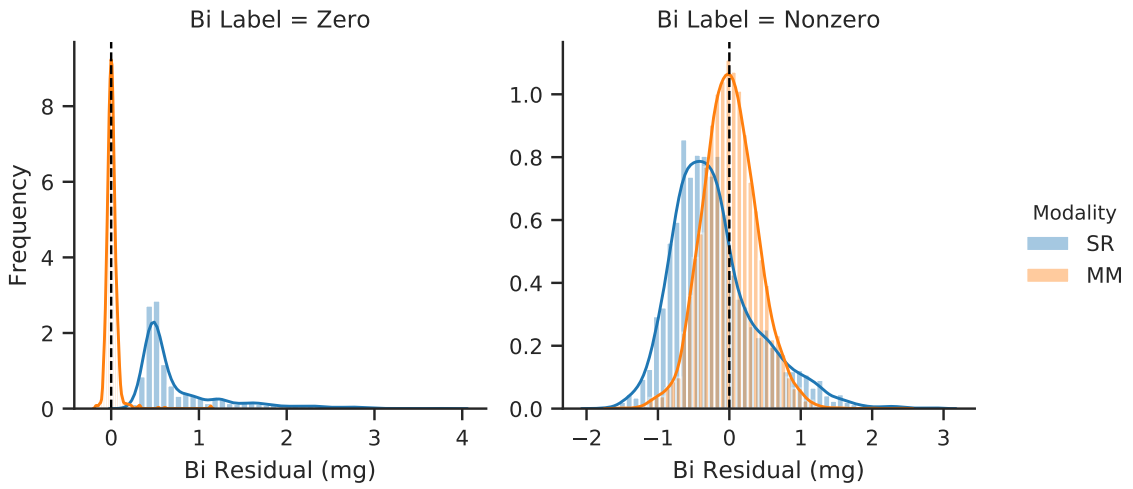

Figure 5.30: Bi residual histograms for SR and MM models on EL4 dataset separated by zero and nonzero Bi labels.

Nonzero Labels. The mean of the residuals for SR was –0.209 mg (99% CI [–0.237 mg, –0.180 mg]) and for MM was 0.00862 mg (99% CI [–0.00946 mg, 0.0267 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test (t(5084.685) = –16.618, $p < 0.001$) and a Mann-Whitney

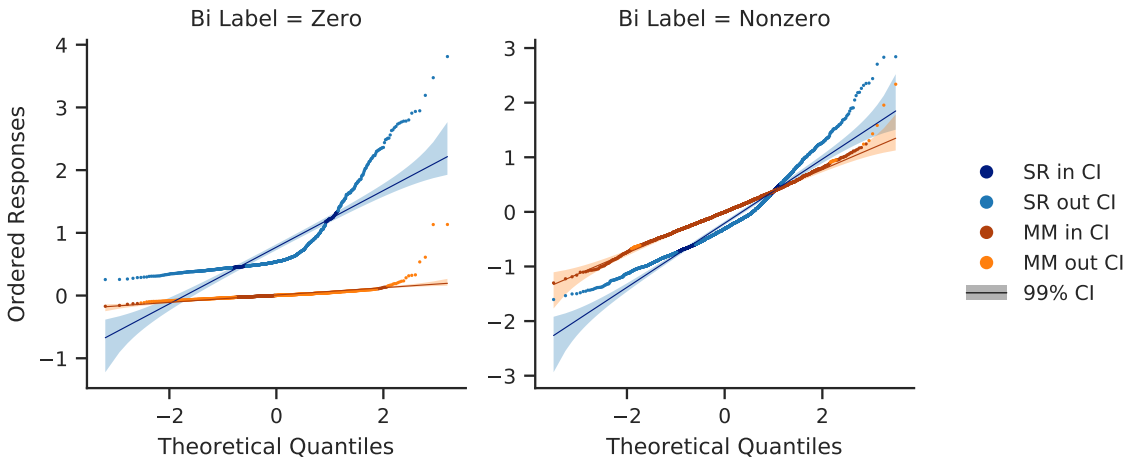

Figure 5.31: Probability plots for SR and MM models on Bi in EL4 dataset.

U-test (U(3000) = 3022439.5, $p < 0.001$). The standard deviations of the residuals were 0.604 mg (99% CI [0.584 mg, 0.625 mg]) for SR and 0.384 mg (99% CI [0.372 mg, 0.397 mg]) for MM. There was statistically significant difference of variances as determined by a Levene test for equal variances (W(1, 5999) = 304.467, $p < 0.001$).

Zero Labels. The mean of the residuals for SR was 0.771 mg (99% CI [0.727 mg, 0.814 mg]) and for MM was 0.00804 mg (99% CI [00179 mg, 0.0143 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test (t(1040.332) = 44.841, $p < 0.001$) and a Mann-Whitney U-test (U(1000) = 2810, $p < 0.001$). The standard deviations of the residuals were 0.532 mg (99% CI [0.503 mg, 0.564 mg]) for SR and 0.0765 mg (99% CI [0.0724 mg, 0.0812 mg]) for MM. There was statistically significant difference of variances as determined by a Levene test for equal variances (W(1, 1999) = 330.326, $p < 0.001$).

Performance of SR and MM Models on O

The O residuals are shown in Figure 5.32 and are separated into examples where the O label was zero or nonzero. Using a Shapiro-Wilk test of normality, we rejected the normality hypothesis for the nonzero O labels (SR: W(3000) = 0.944, $p < 0.001$; MM: W(3000) = 0.986, $p < 0.001$; XRF: W(3000) = 0.977, $p < 0.001$) and for the zero O labels (SR:

W(1000) = 0.824, $p < 0.001$; MM: W(1000) = 0.643, $p < 0.001$; XRF: W(1000) = 0.916, $p < 0.001$). The normal probability plots for O in Figure 5.33 had similar behavior observed for Bi.

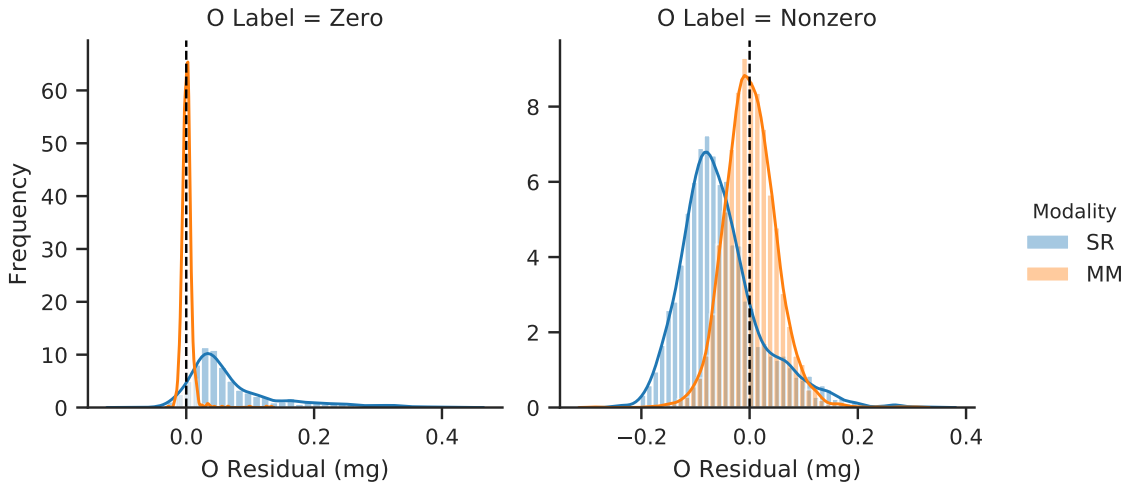

Figure 5.32: O residual histograms for SR and MM models on EL4 dataset separated by zero and nonzero O labels.

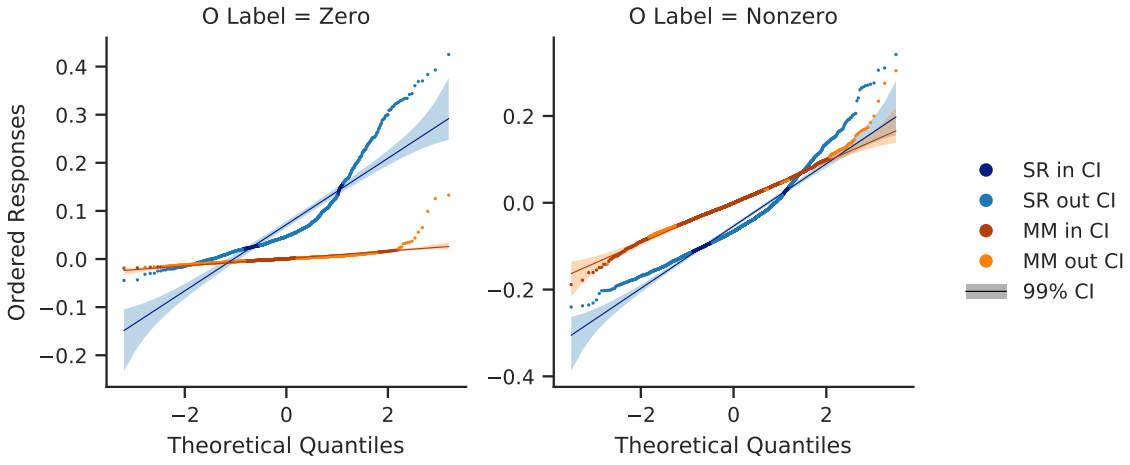

Figure 5.33: Probability plots for SR and MM models on O in EL4 dataset.

Nonzero Labels. The mean of the residuals for SR was –0.0538 mg (99% CI [–0.0573 mg, –0.0504 mg]) and for MM was 0.00118 mg (99% CI [–0.00105 mg, 0.00342 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test (t(5112.643) = –34.278, $p < 0.001$) and a

Mann-Whitney U-test (U(3000) = 2059525.5, $p < 0.001$). The standard deviations of the residuals were 0.0740 mg (99% CI [0.0716 mg, 0.0765 mg]) for SR and 0.0475 mg (99% CI [0.0460 mg, 0.0491 mg]) for MM. There was statistically significant difference of variances as determined by a Levene test for equal variances (W(1, 5999) = 279.983, $p < 0.001$).

Zero Labels. The mean of the residuals for SR was 0.0714 mg (99% CI [0.0652 mg, 0.0777 mg]) and for MM was 0.00108 mg (99% CI [0.000278 mg, 0.00188 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test (t(1032.120) = 29.021, $p < 0.001$) and a Mann-Whitney U-test (U(1000) = 81014.5, $p < 0.001$). The standard deviations of the residuals were 0.0760 mg (99% CI [0.0719 mg, 0.0806 mg]) for SR and 0.00979 mg (99% CI [0.00925 mg, 0.0104 mg]) for MM. There was a statistically significant difference of variances as determined by a Levene test for equal variances (W(1, 1999) = 511.674, $p < 0.001$).

Performance of SR and MM Models on Gd

The Gd residuals are shown in Figure 5.34 and are separated into examples where the Gd label was zero or nonzero. Using a Shapiro-Wilk test of normality, we accepted the normality hypothesis from SR (SR: W(1000) = 0.995, p = 0.00156). We rejected the normality hypothesis from MM and XRF for nonzero Gd labels (SR: W(1000) = 0.983, $p < 0.001$; XRF: W(1000) = 0.988, $p < 0.001$) and from all modalities for zero Gd labels (SR: W(3000) = 0.993, $p < 0.001$; MM: W(3000) = 0.802, $p < 0.001$; XRF: W(3000) = 0.357, $p < 0.001$). The normal probability plots in Figure 5.35 show good agreement to the normality fit with small, long-tail departures for nonzero Gd labels. The departures from normality for zero Gd labels were larger in the tails than the nonzero data.

Nonzero Labels. The mean of the residuals for SR was 0.0202 mg (99% CI [0.00611 mg, 0.0342 mg]) and for MM was 0.0000148 mg (99% CI [–0.0216 mg, 0.0216 mg]). There was no statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test (t(1715.987) = 2.018, p = 0.0438) and a Mann-

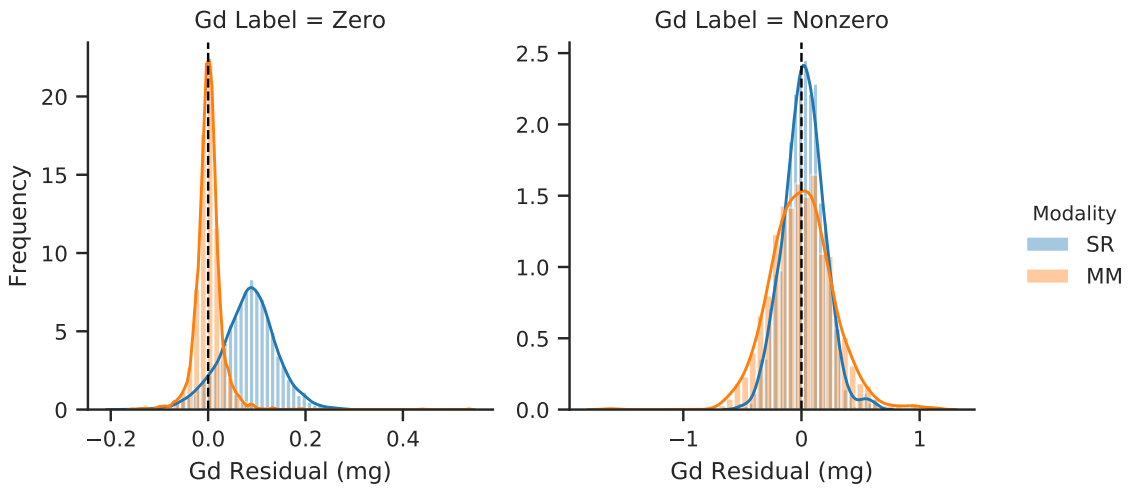

Figure 5.34: Gd residual histograms for SR and MM models on EL4 dataset separated by zero and nonzero Gd labels.

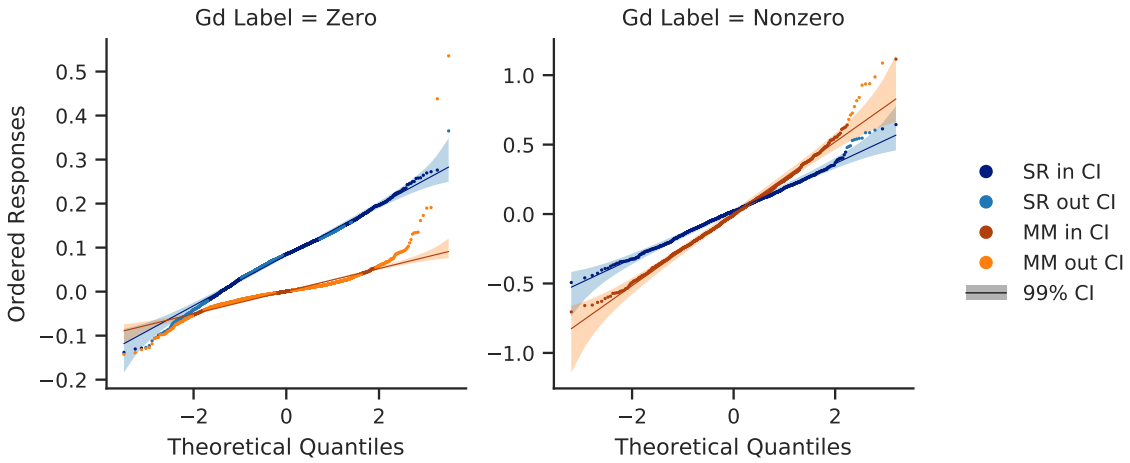

Figure 5.35: Probability plots for SR and MM models on Gd in EL4 dataset.

Whitney U-test (U(1000) = 467021, p = 0.00533). The standard deviations of the residuals were 0.172 mg (99% CI [0.163 mg, 0.183 mg]) for SR and 0.265 mg (99% CI [0.250 mg, 0.281 mg]) for MM. There was a statistically significant difference of variances as determined by a Levene test for equal variances (W(1, 1999) = 122.377, $p < 0.001$).

Zero Labels. The mean of the residuals for SR was 0.0823 mg (99% CI [0.0796 mg, 0.0850 mg]) and for MM was 0.000948 mg (99% CI [–0.000411 mg, 0.00231 mg]). There was a statistically significant difference between the residual means from the two modalities as determined by a Welch's t-test (t(4416.265) = 69.161, $p < 0.001$) and a Mann-Whitney

U-test (U(3000) = 878216, $p < 0.001$). The standard deviations of the residuals were 0.0576 mg (99% CI [0.0557 mg, 0.0596 mg]) for SR and 0.0289 mg (99% CI [0.0279 mg, 0.0299 mg]) for MM. There was a statistically significant difference of variances as determined by a Levene test for equal variances (W(1, 5999) = 1103.047, $p < 0.001$).

5.4 Discussion

The quality of the models were evaluated using numerous metrics. A two-sample K-S goodness-of-fit ($\alpha = 0.001$) was applied to each element, and all models passed except the XRF model on the EL1 dataset (on W). This test does not guarantee that the model fits the data well. Table 5.1 compares the SR and MM models on the basis of model sufficiency, homoscedasticity of residuals, normality of residuals, and polarity of bias for both zero and nonzero cases of label mass.

Scatter plots of the residual vs. label mass were used to assess the sufficiency of the functional part of the model, i.e. whether the model captured the underlying deterministic function. A model whose residuals were strongly correlated with the label mass were determined to be misspecified. Two possible causes for misspecification are an inappropriate functional model or the absence of relevant variables. Both SR and MM were sufficient for all elements on the EL1 and EL2 datasets. The SR models on the datasets EL3 and EL4 were generally insufficient, and the MM models were consistently sufficient on EL3 and EL4. While both SR and MM performed well on EL1 and EL2 by this measure, MM demonstrated improved performance on EL3 and EL4 over SR.

The same scatter plots were used to assess the homoscedasticity of the residuals, i.e. whether the residuals had constant variation across label masses. Both modalities had nearly all homoscedastic residuals. Many cases where SR showed non-constant residuals showed larger variation where the label was zero, but these were not included in the measure. The cases with a visible change in residual variation from SR were improved in MM. By this measure, MM showed marginally better performance.

Table 5.1: Comparison of Model Validations

	Modality		Suff. Model	Homo. Resid.	Norm. Resid.	Bias Polarity	
						label > 0	label = 0
EL1							
	W	SR	✓	✓	✓	insig.	
		MM	✓	✓	✓	insig.	
EL2							
	Bi	SR	✓	×	×	insig.	
		MM	✓	✓	✓	insig.	
	O	SR	✓	✓	×	insig.	
		MM	✓	✓	✓	insig.	
EL3							
	W	SR	×	×	×	$-$	$+$
		MM	✓	✓	×	insig.	$+$
	Bi	SR	×	✓	×	$-$	$+$
		MM	✓	✓	×	insig.	$+$
	O	SR	×	✓	×	$-$	$+$
		MM	✓	✓	×	insig.	$+$
EL4							
	W	SR	×	×	×	$-$	$+$
		MM	✓	✓	×	insig.	$+$
	Bi	SR	×	✓	×	$-$	$+$
		MM	✓	✓	×	insig.	$+$
	O	SR	×	✓	×	$-$	$+$
		MM	✓	✓	×	insig.	$+$
	Gd	SR	✓	✓	×	$+$	$+$
		MM	✓	✓	×	insig.	insig.

The bias polarity was determined using the residual mean and 99% CI. The polarity was positive if the mean was positive and 0 was not within the CI, and vice versa for negative. The bias was considered insignificant if the CI included 0. We separated the analysis into nonzero and zero label masses. EL1 and EL2 showed no significant bias from either modality. EL3 and EL4 consistently had a negative bias on nonzero label masses from SR and no bias from MM aside from Gd which had a positive bias from SR. Alternatively, there was a positive bias on zero label masses from SR and MM.

We directly compared and selected the better performing models in Table 5.2. The sample means and standard deviations of the residuals were determined for each modality on nonzero and zero mass labels. We tested the hypothesis of equal means between modalities with a Welch's t-test (α = 0.001), and the hypothesis of equal variance was tested with a Levene test (α = 0.001). In cases where we failed to reject a hypothesis, the higher performing model was reported as "None." If the t-test null hypothesis was rejected, the higher performing model on residual mean was selected as

$$m_{\text{best}} = \underset{m \in \{\text{SR,MM}\}}{\arg\min} \; |\bar{x}(m)| \tag{5.2}$$

where m is the model modality, and $\bar{x}(m)$ is the sample mean of the residual on model m. If the Levene test null hypothesis was rejected, the higher performing model on residual standard deviation was selected as

$$m_{\text{best}} = \underset{m \in \{\text{SR,MM}\}}{\arg\min} \; s(m) \tag{5.3}$$

where $s(m)$ is the sample standard deviation of the residual on model m. For the EL1 and EL2 datasets, the absolute residual means were not significantly different between SR and MM. The model with higher performance on residual standard deviation depended on the element: SR performing better on W and Bi and MM performing better on O. For the EL3 and EL4 datasets, the MM model had a smaller absolute residual mean and a smaller

Table 5.2: Higher Performing Modality Based on Residual Mean and Std. Dev.

		Selection Criteria			
		Residual Mean		Residual Std. Dev.	
		label > 0	label = 0	label > 0	label = 0
EL1					
	W	None		SR	
EL2					
	Bi	None		SR	
	O	None		MM	
EL3					
	W	MM	MM	MM	MM
	Bi	MM	MM	None	MM
	O	MM	MM	MM	MM
EL4					
	W	MM	MM	MM	MM
	Bi	MM	MM	MM	MM
	O	MM	MM	MM	MM
	Gd	None	MM	SR	MM

residual standard deviation for nearly all elements.

We quantified each model's overall performance on each dataset {EL1, EL2, EL3, EL4}, modality {SR, MM, XRF}, and element {W, Bi, O, Gd} using the coefficient of variation CV(RMSD) in Table 5.3. The root mean square deviation (RSMD) is defined as

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{5.4}$$

where y_i is the label mass on the example i^{th}, $\hat{y}_i$ is the model-predicted mass, and n is the number of test examples. The CV(RSMD) is defined as

$$\text{CV(RSMD)} = \frac{\text{RMSD}}{\bar{y}}. \tag{5.5}$$

We used Miller's confidence interval for CV, defined as

$$\text{CI}_{\text{Miller}} = \text{CV} \pm Z_{1-\alpha/2}\sqrt{\frac{\text{CV}^2}{n-1}\left(0.5 + \text{CV}^2\right)}. \tag{5.6}$$

Table 5.3: Coefficient of Variation Across All Modalities, All Data

	Modality	CV(RSMD) in % (α = 0.01)			
		W	Bi	O	Gd
EL1					
	SR	$0.310\% \pm 0.007\%$	–	–	–
	XRF	$5.289\% \pm 0.119\%$	–	–	–
	MM	$0.342\% \pm 0.008\%$	–	–	–
EL2					
	SR	–	$0.382\% \pm 0.009\%$	$0.733\% \pm 0.016\%$	–
	XRF	–	$11.468\% \pm 0.260\%$	$11.518\% \pm 0.261\%$	–
	MM	–	$0.456\% \pm 0.010\%$	$0.456\% \pm 0.010\%$	–
EL3					
	SR	$0.513\% \pm 0.007\%$	$0.549\% \pm 0.007\%$	$0.664\% \pm 0.009\%$	–
	XRF	$7.746\% \pm 0.101\%$	$13.876\% \pm 0.183\%$	$13.872\% \pm 0.183\%$	–
	MM	$0.348\% \pm 0.004\%$	$0.387\% \pm 0.005\%$	$0.394\% \pm 0.005\%$	–
EL4					
	SR	$0.868\% \pm 0.010\%$	$0.911\% \pm 0.010\%$	$0.986\% \pm 0.011\%$	$3.610\% \pm 0.040\%$
	XRF	$7.792\% \pm 0.088\%$	$15.453\% \pm 0.177\%$	$14.988\% \pm 0.171\%$	$28.645\% \pm 0.346\%$
	MM	$0.351\% \pm 0.004\%$	$0.421\% \pm 0.005\%$	$0.430\% \pm 0.005\%$	$4.037\% \pm 0.045\%$

The CV(RMSD) was used as a measure of precision across modalities and elements. A lower CV(RMSD) value indicates a more precise model.

The values in Table 5.3 for EL3 and EL4 are artificially inflated because of the inclusion of zero label masses. These labels bring down the mean in a way that depends on the number of zero and nonzero labels. For example, only 25% of EL4 included nonzero Gd labels, while 75% include nonzero W, Bi, and O labels. Table 5.4 shows updated values calculated only with the nonzero vales.

On the single-compound datasets, EL1 and EL2, the CV(RSMD) from the SR models were lower than from the MM models for the heavy metals, Bi and W, and higher for O. The CV(RSMD) values from the SR models were higher on the multi-compound datasets than on the single-compound datasets; however, the CV(RSMD) values from the MM models on the multi-compound datasets were either similar or lower than on the single-compound datasets. They MM values were consistently lower than the SR models on the multi-compound datasets. These results suggest that the SR model is superior for single-

Table 5.4: Coefficient of Variation Across All Modalities, Nonzero Labels

	Modality	CV(RSMD) in % (α = 0.01)			
		W	Bi	O	Gd
EL3					
	SR	$0.345\% \pm 0.005\%$	$0.417\% \pm 0.007\%$	$0.506\% \pm 0.008\%$	–
	XRF	$6.323\% \pm 0.100\%$	$11.352\% \pm 0.182\%$	$11.343\% \pm 0.182\%$	–
	MM	$0.284\% \pm 0.004\%$	$0.314\% \pm 0.005\%$	$0.319\% \pm 0.005\%$	–
EL4					
	SR	$0.685\% \pm 0.009\%$	$0.604\% \pm 0.008\%$	$0.714\% \pm 0.009\%$	$1.297\% \pm 0.029\%$
	XRF	$6.741\% \pm 0.087\%$	$13.390\% \pm 0.176\%$	$12.987\% \pm 0.170\%$	$13.962\% \pm 0.318\%$
	MM	$0.303\% \pm 0.004\%$	$0.362\% \pm 0.005\%$	$0.369\% \pm 0.005\%$	$1.983\% \pm 0.044\%$

compound powders, and the MM model is superior for multi-compound powders. For W, Bi, and O, the SR values were consistently $<1\%$, the XRF values were $<15\%$, and the MM values were $<0.5\%$.

The Gd CV(RSMD) values on EL4 were significantly higher than other elements. The SR and MM models had a larger difference in values for Gd from other elements, from 0.3%–0.7% to 1%–2%. The XRF model had a smaller relative difference, from 7%–13% to $\sim$14%.

The results have shown that a deep learning model trained on SR data can perform high-precision ($<1\%$) elemental mass quantification under geometric uncertainty and from noisy (10 min @ 40 μA) acquisitions. In cases with only one compound, *i.e.* a single task, there was little benefit gained from an additional XRF sensor. However, a target composed of a random powder or mixture of powders was more challenging for the SR model, even with a small set of powder types. A fused model combining both SR and XRF data out-performed the model trained only on SR in most cases.

The models trained solely on XRF data performed poorly on mass quantification which suggests the data carried little predictive power. The uncertainty of the target geometry and composition limit the amount of information available to the XRF sensor. The amount of material is degenerate with a given signal because multiple targets with different compositions can produce similar XRF signals. The XRF sensor benefits from close proximity to

the target and measures the X-rays outside the primary beam. It primarily collects signal induced near the surface of the target on the sides closest to the detector face. Although the quantitative information is poor, the signal can be used to identify which elements are present in the target. We speculate that this qualitative information helps simplify the mass quantification task when fused with SR data. This explains the superior performance of the MM model on multi-composition tasks where the powder composition is unknown.

The fusion model used in this work first trained separate models on SR and XRF. The outputs of the final hidden layers were extracted as features which were then combined to form a new dataset. The fused dataset was used to train a model from the combined, extracted features. This approach uses representation learning to determine the most relevant features from each input stream. This methodology benefits from simplicity during training because each modality can be fine tuned individually. Separate training also allows for deeper unimodal networks when GPU capacity is limited. Two different approaches may provide a more fruitful data fusion. First, an autoencoder may provide better features for data fusion. As opposed to our greedy approach which assumes the best features should be able to directly perform the task, an autoencoder can learn a richer representation of the dataset. The features from autoencoders would have the potential for a more meaningful fusion of the datasets. Second, the early fusion of the networks into a single network would attempt to learn a joint representation. The approach can exploit low-level correlations between modalities that are ignored with independent feature extraction.

The poor performance on quantifying Gd mass has a number of contributing factors. As discussed earlier, there was an imbalance of labels. Only 25% of examples contained nonzero Gd labels in EL4. Further, the training set contained examples with single compounds of only W and only Bi_2O_3, but Gd_2O_3 was only present in mixtures of W + Bi_2O_3 + Gd_2O_3. The features of Gd were, therefore, always presented in combination with the W and Bi features. Learning to predict Gd mass under these conditions is challenging because the learning steps are skewed in favor of optimizing the more abundant and dominant

features.

The Gd prediction task also faced physical obstacles. First, the Gd content was randomly sampled to be between 5% and 10% by mass. The approximate upper bounds for each element's total mass in a single sample were: $\sim$380 mg of W, $\sim$230 mg of Bi, $\sim$27 mg of O, and $\sim$24 mg of Gd. While O and Gd had similar mass ranges, the amount of O was highly correlated to the amount of Bi, but Gd was independent of the W and Bi content. The signal induced by Gd was consequently limited.

The XRF signal from Gd is clearly visible in Figure 5.4 between 40 and 50 keV, but the majority of the quantitative information comes from the SR data which measure transmission. In addition to the amount of each element, the induced SR signal in a pixel depends on the X-ray energy distribution and the mass attenuation coefficients of the elements in the target, shown in Figure 5.36. The K-edge jumps present in mass attenuation coefficients are responsible the element-specific information. Although the Gd K-edge is larger than either Bi or W, the X-ray beam intensity around the corresponding energy is relatively small due to an Al filter. This filter configuration was designed to optimized the signal from high-Z targets. The low beam intensity around the Gd K-edge and the small quantity of Gd result in a small signal-to-noise ratio for Gd.

This methodology has the potential for better performance on Gd which it may achieve by addressing some of the obstacles described above. First, it is prudent to use a beam with an energy distribution that has higher intensity between 40 and 50 keV. A thinner Al filter would accomplish this goal. Second, a larger dataset that balances the Gd labels with the other elements present. The additional data would include Gd_2O_3 alone as well as more evenly mixed mixtures of Gd_2O_3 with the other compounds of interest. Although a doping scheme may be of interest, the additional data outside this domain may provide information to the network.

A lingering question around this work stems from the use of simulated data for training: how well can these networks perform on experimental data? There were several efforts

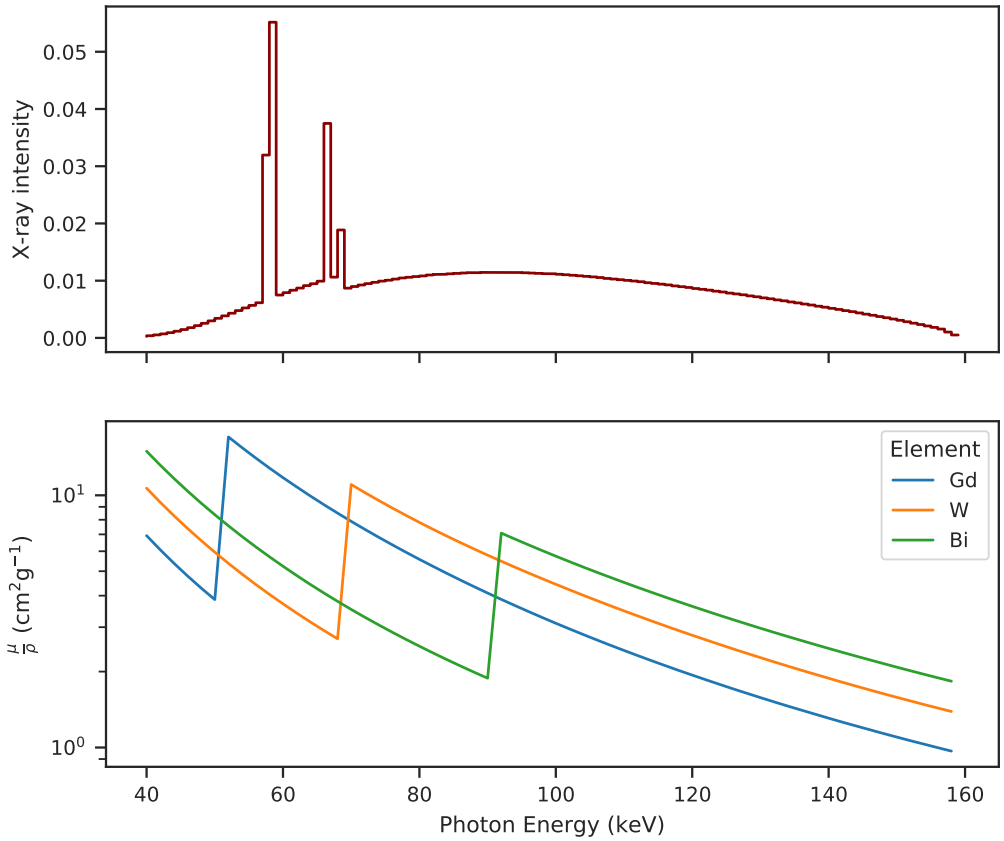

Figure 5.36: The energy distribution of X-rays impinging on the target is shown with the mass attenuation coefficients of Gd, W, and Bi. The beam has passed through a 3-cm Al filter to optimize the signal from high-Z targets.

made to make the simulated data realistic. The randomized geometry and powder density distributions emulated some of the variability found in experimental data. Additionally, the validated detector response function and Poisson noise were selected to closely match observation in the laboratory. There are still features that remain uncaptured in simulation. These features constitute a domain shift between simulation and reality. Chapter 6 attempts to address the discrepancy between these domains using a deep learning approach to domain adaptation.

CHAPTER 6

DEEP DOMAIN ADAPTATION BETWEEN SIMULATION AND EXPERIMENT

6.1 Introduction

Deep learning provides an alternative to algorithmic approaches for quantitative spectral X-ray radiography. The expressive power of deep neural networks is advantageous for applications where the imaging system may significantly deviate from the physics model and assumptions used in an algorithmic approach. Traditional approaches often require high certainty on aspects of the system like geometric parameters, the detector's response and calibration, and the X-ray beam's spatial and energy distributions. Incorporating systematic uncertainty in the training set yields a deep learning model capable of quantification from data subject to that systematic uncertainty. Consequently, the mitigation measures used in traditional approaches can be relaxed, allowing systems to be more cost-effective and simpler to operate.

Deep neural networks typically require large datasets to inductively learn a mapping from the data to a desired output. The data, however, may be prohibitive to obtain. For example, spectral X-ray radiography data obtained by collecting frames of pixel-wise energy deposition face challenges in acquisition and processing. Low currents are necessary to achieve optimal performance resulting in long acquisition times. In addition, the raw data files are large and are cumbersome to manage and process. In this and many cases, collecting sufficient experimental data for deep learning is impractical.

Applications where data are generated through physical processes, like spectral X-ray radiography, can often use simulation to synthesize data. Large-scale computing has made simulated data attractive for generating large training datasets for deep learning. However, deep learning models trained in a simulated domain, the source domain, struggle to transfer

the gained knowledge when applied to an experimental domain, the target domain. A large domain shift between experiment and simulation is a major drawback of using simulated training data.

Domain adaptation has been used to address the domain shift between simulated and real-world data for a variety of tasks [113–115]. Domain adaptation is a type of transfer learning that learns to perform a task in a target domain by learning from labeled data in a different, source domain. Deep learning architectures have been used in deep domain adaptation, contributing rich, hierarchal representations and the flexibility to perform on a wide variety of domains. Recent advances have focused on adversarial-based and representation-based learning approaches. Wang *et al.* [116] provides a survey and taxonomy of deep domain adaptation, focusing on visual domains. Wilson *et al.* [117] reviewed the state-of-the-art of unsupervised deep domain adaptation where there are no labeled data from the target domain.

In this work, we used deep domain adaptation to address the gap between simulated and experimental spectral radiography data. We focused on the task of mass quantification of Bi_2O_3 samples. Using a large, labeled dataset of simulations and a limited number of experimental measurements, we first performed reverse domain adaptation. This step mapped experimental data to the simulation domain using adversarial training. We then used a regression model trained on simulation data to determine the mass on translated experimental data.

6.2 Related Work

Simulating data has many practical advantages, facilitating applications in deep learning that would otherwise be prohibitive. For example, deep reinforcement learning has been of recent interest for robotic visual tasks where an agent learns a task by interacting with an environment. Agents are often trained in simulated environments, but they must address the gap between reality and simulation when transferring to the real world. There has been

a large effort to solve the *reality gap* problem with domain adaptation for robotic tasks such as manipulation [113, 118–124], navigation [125–127], and object detection [128, 129].

Hyperspectral imaging for remote sensing produces 3D spatio-spectral data similar to spectral radiography data. Spectral shifts in hyperspectral images due to environmental variations can cause a shift in the domain, and training on one domain may not be applicable to a different, test domain. Domain adaptation has been used to address this domain shift for classification problems in hyperspectral imaging [114, 120, 130, 131].

Domain adaptation has also been widely explored in medical imaging. The applications include synthesis [115], reconstruction [132], segmentation [133], and classification [134] on a range of modalities such as MRI [135], CT [136], and histopathology [137]. More comprehensive discussions on this topic are available in review articles [138, 139].

Generative adversarial networks (GANs) are at the heart of many domain adaptation methods. Goodfellow *et al.* [140] first presented GANs as a method to synthetically generate samples from a target distribution. They used game theory to construct a pair of adversarial networks: a generator and a discriminator, shown in Figure 6.1. The task of the generator is to produce samples from a noise vector, and the task of the discriminator is to distinguish samples produced by the generator from samples drawn from training data. During training, the goal of the generator is to fool the discriminator which is the contrary of the discriminator's goal.

Conditional GANs (cGANs) were an early variation that allowed the model to be conditioned on external information such as class label presented first on MNIST data [141] then on face image data [142]. Image-to-image translation later used cGANs conditioned on images from one domain to generate images in another domain by adapting the generator network structure. For example, variational autoencoders (VAE) [143] are a type of generative network that were proposed as the generator conditioned on images, known as VAE-GANs [144]. Figure 6.2 shows three image-to-image translation frameworks. Isola *et al.* [145] presented pix2pix which uses a U-Net generator [146] and pairs of images from

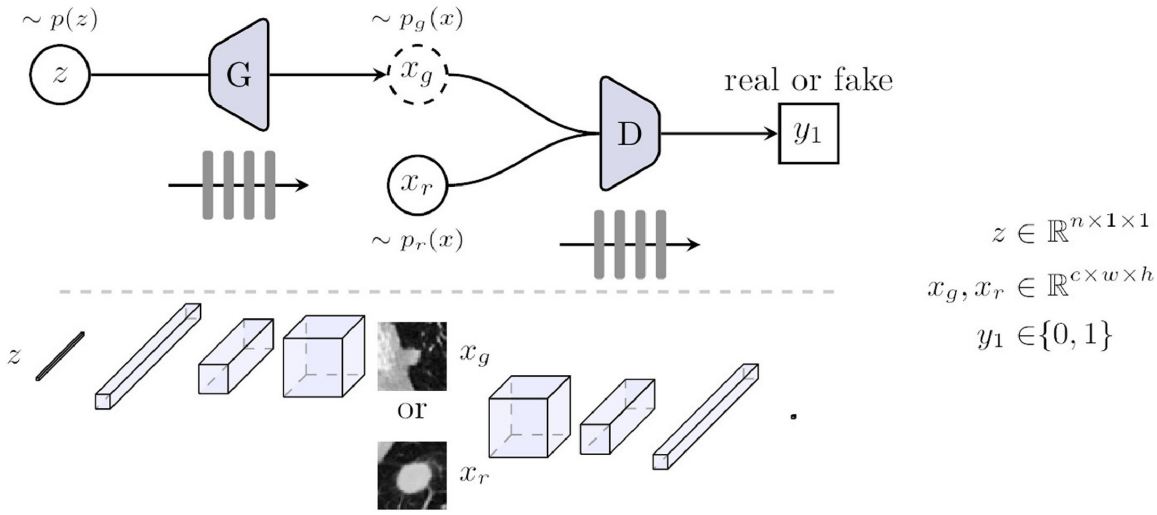

Figure 6.1: A standard GAN architecture consists of a generator network (G) and a discriminator network (D). A random noise vector (z) feeds into the generator which attempts to output a realistic sample. Real data and generated data feed into the discriminator which attempts to discriminate real from generated data [138].[1]

different domains.

In many cases, paired images are unavailable, and a different architecture is necessary. Zhu *et al.* [147] developed CycleGAN for unpaired image-to-image translation by learning two mapping functions, one from domain A to domain B and another from domain B to domain A. The training was constrained by a cycle-consistency loss [148] where an input passed through the first mapping then the second mapping should return that same original input. Liu *et al.* [149] used two VAE-GANs with a shared latent space to formulate the unsupervised image-to-image translation (UNIT) framework for unpaired image domains.

In unsupervised domain adaptation using domain mapping, the source domain is typically translated to the target domain. Subsequently, the classification or regression task is trained on the translated source data with the corresponding labels and tested on the target data. The classifier or regressor and the domain mapping may be trained independently or jointly. Shrivastava *et al.* [150] proposed the SimGAN framework that "refined" simulated image data for gaze estimation to look like realistic images, and the regression network

[1]Reprinted from Medical Image Analysis, Vol. 58, Xin Yi, Ekta Walia, Paul Babyn, "Generative adversarial network in medical imaging: A review," pp. 101552, Copyright (2019), with permission from Elsevier.

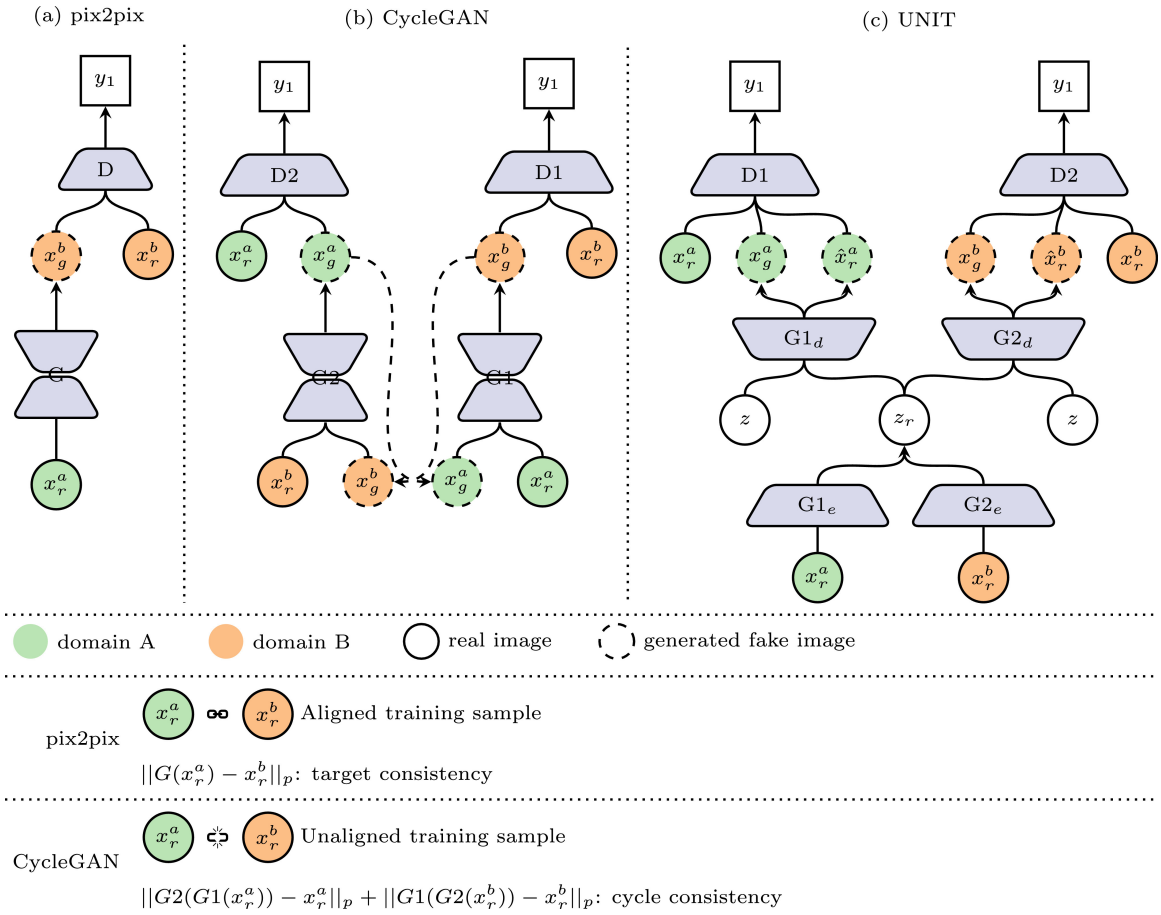

Figure 6.2: Adversarial domain adaptation extends the idea of cGANs by conditioning the generator on one domain to generate data from another domain. Image-to-image translation is an example of the domain mapping approach to unsupervised domain adaptation [138].[2]

was trained independently. Mahmood *et al.* [151] used a similar approach for depth estimation, but they instead mapped the unlabeled target domain of real images to the labeled source domain of simulated images. Both used adversarial architectures where the translator acted as a GAN generator which fed into a discriminator network. In order to preserve the labels during translation, both included a self-regularization loss that minimized the L1 norm between the simulated images and the translated, simulated images. Bousmalis *et al.* [152] shared a similar architecture, but they replaced the L1 norm self-regularization with a jointly trained classifier.

[2]Reprinted from Medical Image Analysis, Vol. 58, Xin Yi, Ekta Walia, Paul Babyn, "Generative adversarial network in medical imaging: A review," pp. 101552, Copyright (2019), with permission from Elsevier.

Table 6.1: Masses of Bi_2O_3 powder samples.

ID#	Bi_2O_3 Mass
1	0.1604 g
2	0.1264 g
3	0.1343 g
4	0.2136 g
5	0.2016 g
6	0.1444 g
7	0.1497 g
8	0.1742 g
9	0.1672 g
10	0.1740 g
11	0.1723 g
12	0.1681 g
13	0.1866 g
Avg.	0.1671 g
Std. Dev.	0.0239 g

6.3 Materials and Methods

6.3.1 Experimental and Simulated Datasets

The powder samples of Bi_2O_3 were prepared using the procedure described in Subsection 3.3.3. Initially, 13 Bi_2O_3 samples were prepared. The number of cuvettes and samples that we were able to prepare was limited by time. Table 6.1 shows the masses of the powder in each of the final samples.

Each of the samples was measured multiple times to efficiently increase the size of the dataset. There were two orientations used for a sample. The lid of a forward-oriented (F) sample faced the X-ray source, and the lid of a reverse-oriented (R) sample faced the X-ray detector. We cycled through the samples and alternated orientations between acquisitions. Table 6.2 shows the number of 10-min acquisitions taken for each sample in both orienta-

Table 6.2: Number of 10-min acquisitions with sample oriented forward (F) and reversed (R).

	Bi_2O_3	
ID#	F	R
1	4	2
2	4	2
3	4	2
4	4	2
5	4	2
6	4	2
7	4	2
8	4	2
9	4	2
10	2	1
11	2	1
12	2	1
13	2	1
Total	40	22
Target Total	62	

tions. Additionally, we took five 10-min flatfield measurements that were aggregated using the method described in Subsection 3.3.6.

The Monte Carlo simulations follow the description in Section 3.2, including geometric and density randomization. Assuming that each powder occupies a 1 cm $\times$ 1 cm $\times$ 1 mm volume, we estimated the powder densities from the masses in Table 6.1. The average $\bar{\rho}$ and population standard deviation s_ρ of the densities were used to determine the distribution of densities across the simulations. The mean density used for the i^{th} simulation was sampled as $\mu_{\rho,i} \sim \mathcal{N}(\bar{\rho}, s_\rho^2)$. We generated 5000 simulations for Bi_2O_3 that were designated as training sets. We also generated an additional 1000 simulations for validation and 1000 for testing.

The simulated and experimental data were processed in similar ways and both ended with the same overall structure. After being parsed (detailed in Subsection 3.3.5), the experimental data y_{exp} were rebinned into 1 keV bins, and the scaled flatfield $y_{0,\text{exp}}$ was determined for each transmission measurement (detailed in Subsection 3.3.6). The simulated transmission data y_{sim} were transformed by the HEXITEC response function (detailed in Subsection 3.1.2), scaled to match the intensity of a 10-min acquisition, and blurred with Poisson noise. The simulated flatfield $y_{0,\text{sim}}$ was also transformed by the response function and scaled in intensity, but no noise was added.

Energy thresholds were applied to the transmission and flatfield data from both simulation and experiment, keeping energies between 60 keV and 140 keV. The transmission data were then transformed using the flatfield data into

$$\begin{aligned} d_{\text{exp}} &= 1 - \frac{y_{\text{exp}}}{y_{0,\text{exp}}} \\ d_{\text{sim}} &= 1 - \frac{y_{\text{sim}}}{y_{0,\text{sim}}}. \end{aligned} \tag{6.1}$$

For any case where y_0 = 0, y/y_0 was set to 1. The transformed data were then confined to $0 \leq d \leq 1$ by setting $d = 0$ where $d < 0$ and $d = 1$ where $d > 1$. Next, they were scaled by a factor of 2 and translated by –1 such that $-1 \leq d \leq 1$. Lastly, we augmented the experimental data d_{exp} using the 8 unique transformations from horizontal flips and 90° rotations, yielding 496 experimental Bi_2O_3 examples.

We visualize the two Bi_2O_3 datasets (simulated and experimental) before training. We show the energy-averaged 2D image (Figure 6.3) and the energy spectrum from pixel (40, 40) (Figure 6.4) for a number of samples from each dataset before the final scaling and translation. The color scale of the images and the y-scale of the spectra are both from 0 to 1.

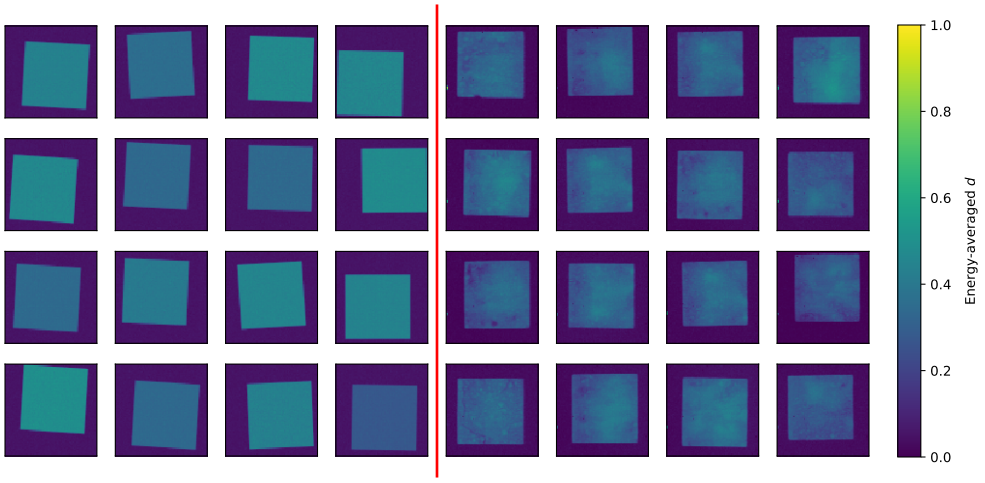

Figure 6.3: Comparison between simulated (left) and experimental (right) energy-integrated radiographs.

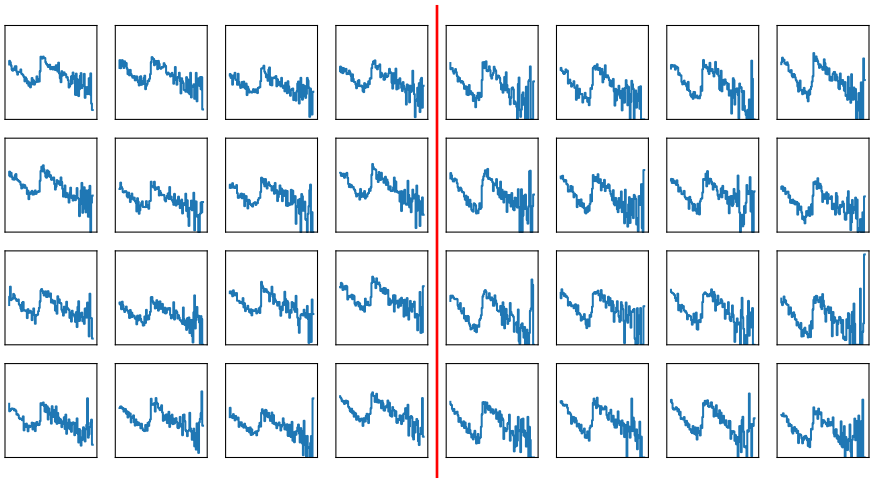

Figure 6.4: Comparison between simulated (left) and experimental (right) energy spectra from the center pixels.

6.3.2 Deep, Unsupervised Domain Adaptation Architecture

The unsupervised domain adaptation of data from one domain $\boldsymbol{x} \sim \mathcal{X}$ to another $\boldsymbol{y} \sim \mathcal{Y}$ comprised two primary network structures: a translator and a discriminator. The translator network produced data $\boldsymbol{t} = T(\boldsymbol{x}; \boldsymbol{\theta}_T)$ translated from $\boldsymbol{x}$ and parameterized by $\boldsymbol{\theta}_T$. The task of the translator was to generate data $\boldsymbol{t}$ that was indistinguishable from data from $\mathcal{Y}$. More broadly, the translator functioned as the generator in a conditional GAN that was conditioned on data from $\mathcal{X}$. The discriminator network $D(\cdot; \boldsymbol{\theta}_D)$ decided whether an input

came from $\mathcal{Y}$ or was translated from $\mathcal{X}$. The parameters for the discriminator network were $\boldsymbol{\theta}_D$.

The translator network used a U-net architecture [146] which has had success in image-to-image translation [153]. The first stage of U-net was an encoder that mapped the input to a low-dimensional feature space that acted as a bottleneck. The downsampling was done using convolutional layers. The second, decoder stage then mapped the feature space to an output with the same dimensions as the original input. Skip connections were formed by concatenating the outputs of the encoder layers to the decoder layers with its corresponding shape. Figure 6.5 depicts the U-net translator network used in this work. Batch normalization was neither applied before the first hidden layer nor after the last hidden layer. The activation functions were leaky ReLu in the encoder and ReLu in the decoder. The selection of architectural features followed guidelines for deep convolutional GANs in [154].

The discriminator network used a 3D convolutional neural network architecture. Figure 6.6 depicts the full network. The convolutional layers reduced the shape using a kernel size of 4 and stride of 2 (or 5 and 1, respectively, for the final convolutional layer). The last hidden layer was followed by a fully connected layer containing a single output and a sigmoid activation function, $\boldsymbol{d}$. We used a leaky ReLu activation function for the remaining (hidden) layers. Just as for the translator network, batch normalization was neither applied to the input nor output.

In addition to the discrimination output, we added a second fully connected layer with a single output unit to the penultimate layer. This output node had no activation function. Instead of a domain classifier, it functioned as an auxiliary regressor that calculated value $\boldsymbol{m}_{\text{aux}}$. The network parameters $\boldsymbol{\theta}_{\text{aux}}$ were nearly all shared with $\boldsymbol{\theta}_D$, but the final layer parameters were exchanged with the parameters from regressor's fully connected layer. The purpose of the regressor was to allow for supervision using sample mass during training. It serves as an estimate for the mass of translated data. We compared the performance with the auxiliary regressor to the completely unsupervised network on the domain adaptation

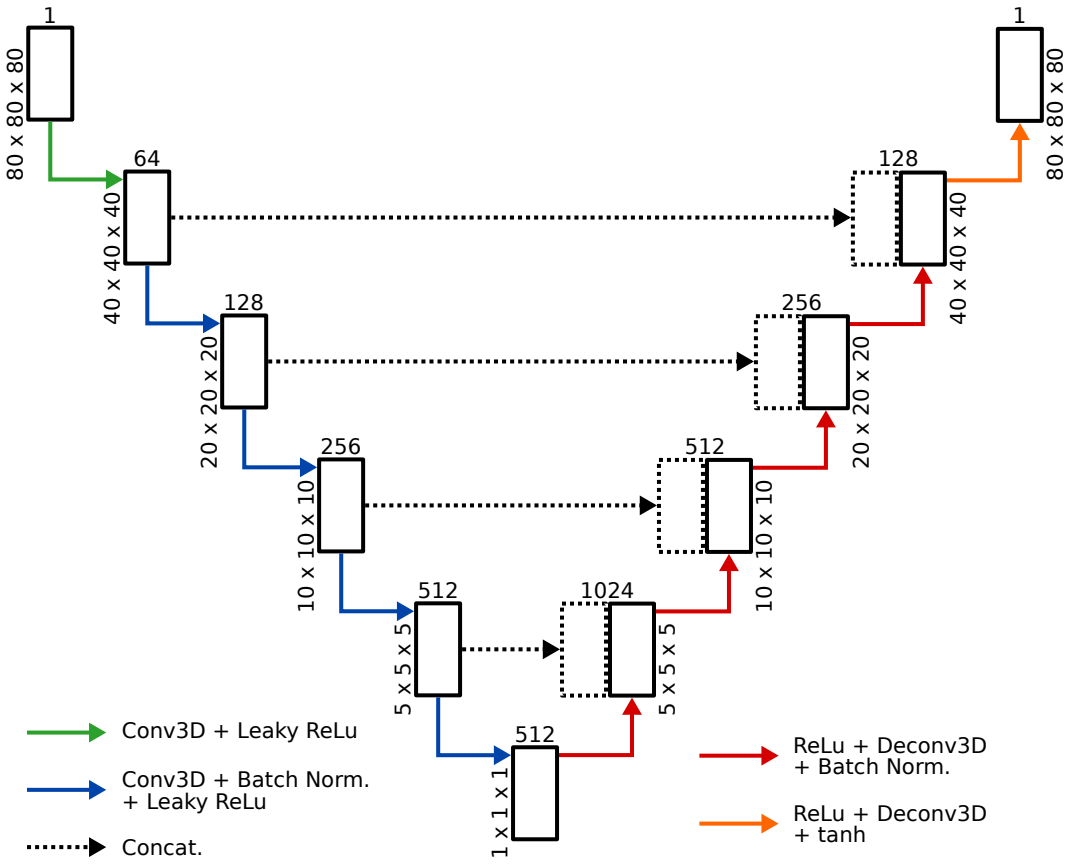

Figure 6.5: The translator network translated data from one domain to the other using a U-net architecture. It had an encoder-decoder structure that forced the data through a low-dimensional bottleneck layer. The five convolutional layers in the first half of the network made up the encoder, and the five deconvolutional layers after the bottleneck formed the decoder. Skip connections bypassed the bottleneck by concatenating encoder layer outputs with their corresponding decoder layers.

task.

6.3.3 Loss Function Formulation and Training

The standard formulation of a GAN is motivated by game theory. The discriminator tries to maximize the log-probability of correctly whether the input came from $\mathcal{Y}$ or was translated from $\mathcal{X}$. The translator, however, tries to minimize the log-probability that the discriminator does not make a mistake. The overall task can be formulated as

$$\min_{T} \max_{D} \mathbb{E}_{\boldsymbol{y} \sim \mathcal{Y}} \log D\left(\boldsymbol{y}\right) + \mathbb{E}_{\boldsymbol{x} \sim \mathcal{X}} \log\left(1 - D\left(T\left(\boldsymbol{x}\right)\right)\right). \tag{6.2}$$

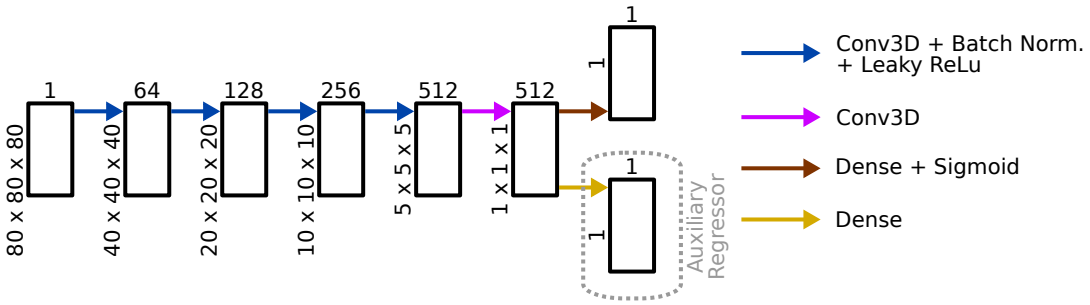

Figure 6.6: The discriminator network architecture primarily consisted of 3D convolutional layers that downsampled the data. After the final convolutional layer, a dense layer and sigmoid activation provide the discrimination output. A value of 1 indicates the discriminator believes the input to be real data, and a value of 0 indicates it believes it came from the translator network.

The optimization is performed by alternating between separate gradient steps for D and T. From Equation 6.2, the D step should maximize

$$\mathbb{E}_{\boldsymbol{y}\sim\mathcal{Y}} \log D\left(\boldsymbol{y}\right) + \mathbb{E}_{\boldsymbol{x}\sim\mathcal{X}} \log\left(1 - D\left(T\left(\boldsymbol{x}\right)\right)\right) \tag{6.3}$$

over $\boldsymbol{\theta}_D$, and the T should minimize

$$\mathbb{E}_{\boldsymbol{x}\sim\mathcal{X}} \log\left(1 - D\left(T\left(\boldsymbol{x}\right)\right)\right). \tag{6.4}$$

Goodfellow *et al.* [140] proposed that the gradient step be reformulated as maximizing

$$\mathbb{E}_{\boldsymbol{x}\sim\mathcal{X}} \log D\left(T\left(\boldsymbol{x}\right)\right) \tag{6.5}$$

to help in early learning stages. The T and D steps are then both maximization steps. By flipping the signs of each loss function, we instead minimized at each step using gradient descent.

The translator training step minimized

$$\mathbb{E}_{\boldsymbol{x}\sim\mathcal{X}} - \log D\left(T\left(\boldsymbol{x}\right)\right) \tag{6.6}$$

over the translator parameters $\boldsymbol{\theta}_T$. The discriminator parameters $\boldsymbol{\theta}_D$ were used to calculate the gradients through the discriminator but were not changed during a T step. The inputs $\boldsymbol{x}$ to the network were all sampled from $\mathcal{X}$, and 50% of the resulting translated data $\boldsymbol{t}$ from a step were added to an experience replay buffer in the form of a queue. The queue added examples to its buffer until full, and new examples added to a full queue randomly replaced existing queue items. More on the use of experience replay for GANs can be found in [150, 155, 156]

The discriminator training step minimized

$$\mathbb{E}_{\boldsymbol{y}\sim\mathcal{Y}} - \log D\left(\boldsymbol{y}\right) + \mathbb{E}_{\boldsymbol{x}\sim\mathcal{X}} - \log\left(1 - D\left(T\left(\boldsymbol{x}\right)\right)\right) \tag{6.7}$$

over the discriminator parameters $\boldsymbol{\theta}_D$ (with $\boldsymbol{\theta}_T$ frozen). The inputs included both $\boldsymbol{y}$ data and translated data from $\boldsymbol{x}$. During a D training set, half of the translated data was generated from sampled x data using the current parameters $\boldsymbol{\theta}_T$, while the other half was randomly sampled from the experience replay buffer.

The auxiliary regressor was included using the label of the input data. The mean squared error between the label and the predicted mass was used as the loss function. This loss was added to both T and D steps and was scaled by parameter λ_{aux}. The parameters of the discriminator were frozen during T training, and vice versa.

6.3.4 Reverse Domain Adaptation

The standard approach to domain mapping for domain adaptation maps the source domain to the target domain. The source domain has labels, and the target domain may or may not have labels. After the mapping, the target-mapped source domain data is used to learn a task so it can be applied directly to target domain data. The simulated data would be the source data that are mapped to the experimental (target) domain. A regression model would then learn from the translated simulated data to determine the mass from experimental data.

Reverse domain adaptation instead translates the experimental data into the simulated

domain. Mahmood *et al.* [151] argued for reversing the flow of domain mapping for labeled synthetic data and unlabeled experimental data in the context of medical imaging. They mapped experimental data to look more like synthetic data. One reason they argued was that synthetic models typically lack some of the complexity and diversity in experimental data. The models also did not generalize across different patients in the original formulation. Our experimental dataset was sparse relative to the simulated dataset which provides an additional challenge. We used a reverse flow for domain adaptation. In the above formulation of the loss functions, the experimental data $\boldsymbol{x}$ was mapped to look like simulated data $\boldsymbol{y}$. The labeled, simulated data was used to learn the regression task.

6.3.5 Supervised Regression Model

Our primary goal for this work was to create a regression model that quantified mass from experimental data. A large, simulated dataset with mass labels would be used for the supervised training of a regression model. We used domain mapping to translate experimental data into the simulated domain. We trained a regression model using only simulated data in the training set. This model was applied to the translated experimental data to predict the mass for that measurement. The regression model used a 3D ResNet-34 architecture. The network outputs contained the elemental Bi and O mass estimates.

6.4 Results

6.4.1 Unsupervised Domain Adaptation

The domain adaptation stage was prone to instabilities due to the adversarial nature of the translator and the discriminator. The accuracy of the discriminator was used as a performance indicator during training. At every training step of the discriminator, we evaluated two accuracy values. The simulation accuracy was the fraction of simulated data inputs that the discriminator labeled as 1, and the translation accuracy was the fraction of the inputs that were translated from experimental data that the discriminator labeled as 0. Figure 6.7

shows these accuracies over the training period. The simulation accuracy was typically higher than the translation accuracy, but stable training had accuracies < 1 on average. Instabilities were reflected in the accuracies typically as both going to 1, either at the beginning of training or after a prolonged training period.

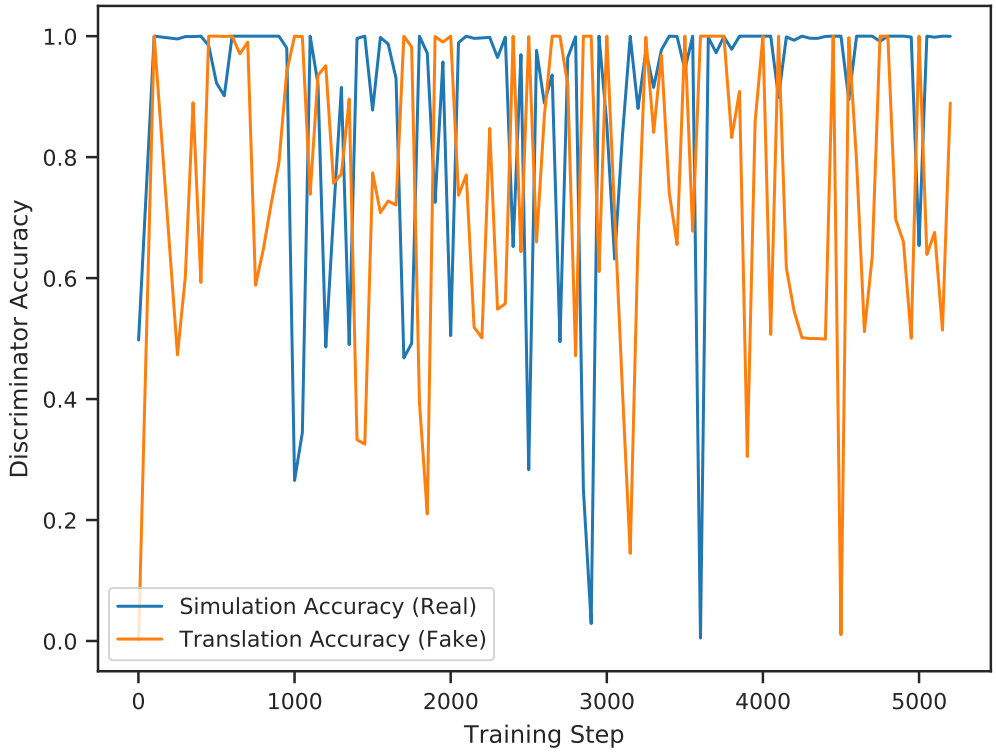

Figure 6.7: The accuracy of the discriminator during discriminator training steps is separated into simulation inputs (blue) and translated experimental inputs (orange). Half of the translated inputs were sampled from the experience replay buffer while the other half came from sampling and translating experimental inputs with the up-to-date translator.

The loss of the networks served as another metric during training. The translation and discrimination losses are shown in Figure 6.8. The translation loss was evaluated during translator training steps and is a measure of how well the translator was able to fool the discriminator. A continually increasing translation loss was indicative of an instability where the translator network failed to converge. This was mitigated by increasing the network capacity or slowing the training rate. Convergence failure was also caused by an experience

replay buffer that was too large. The discriminator loss instead measured how well the discriminator was able to accurately distinguish simulated data from translated experimental data. When the training was stable, the translator loss was higher than the discriminator losses, and none had strong trends beyond the early training steps. Convergence instabilities presented in the discriminator losses by approaching 0.

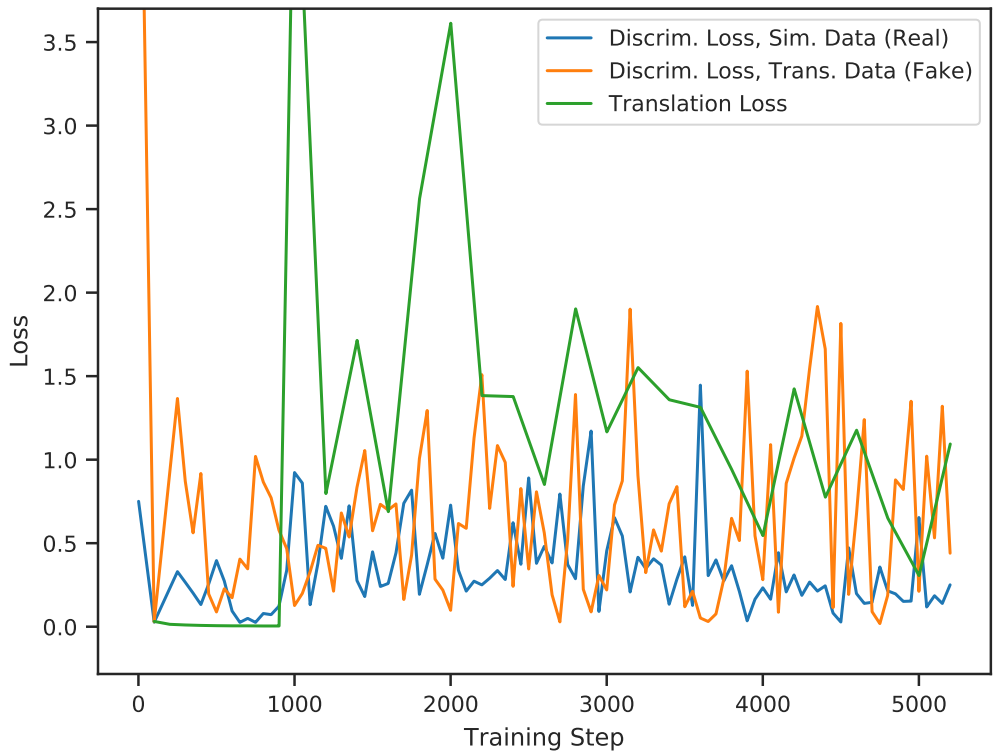

Figure 6.8: The discrimination losses measured the loss of the discriminator during a discriminator training step on simulated input data (blue) and translated experimental data (orange). The translation loss (green) was the loss of the discriminator during a translation step on experimental input data.

The final evaluation metric for the domain adaptation was visual inspection of the translated data, particularly the energy-averaged image data. They provided useful, albeit qualitative, information about the types of instabilities and the overall training progress. Early stages of training consistently produced translated images that were only noise. Low frequency information such as the powder target outline in the image or spectral features

were the first to form. In the case of convergence instability, the translated data did not develop these features. Alternatively, the training could suffer from mode collapse instabilities which were not strongly evident from the accuracy or loss. The translated images from mode collapse were identical for different input data. Mode collapse could be mitigated using a larger experience replay buffer, slowing the training, or rebalancing the training rates between the translator and discriminator. Lastly, the images would often contain artifacts in late training stages as the translator was determining the high frequency information. The artifacts appeared in the form of unrealistic powder sample shapes or strong spatial features apart from the sample's projection. Early stopping prevented the development of these artifacts but resulted in blurrier translated images. Increasing the capacity of the networks, usually the discriminator, helped remove the artifacts. Changing the networks required rebalancing of the training rates.

The images of the 62 experimental measurements are shown before (Original) and after (Translated) translation in Figure 6.9, Figure 6.10, Figure 6.11, and Figure 6.12. The samples are ordered by the experimentally measured mass of the powder. Each powder was measured between 3 and 6 times, each on a different day. All measurements of a sample are shown in a single row. All images share a common color scale. The translated images generally maintained the spatial location and orientation of the sample. The projected areas of the samples in the translated images were more flat, lacking many of the spatial features caused by differences in powder density. This aspect of the translation was expected because simulations also lacked such spatial features. Sample #6 in Figure 6.9, however, shows that some spatial variation of density was maintained through translation. It is unclear from the images whether the total mass was preserved, but the average pixel intensity in the powders' projected areas were higher in the translated images. The multiple translations of the same sample from different measurements, however, appeared to be similar in intensity.

The translated images had noticeable differences from the expected simulated images.

They were generally less sharp, particularly at the edges and corners of the samples' projected areas. This kind of blurring is apparent in the pixel intensities which are much smoother than in the simulation data. The translated images also contain spatial artifacts in the pixels near the image borders (within around 25% of the image width from any edge) in the form of a rectangular frame. When the sample is absent in an area, these pixels are elevated in intensity. When the sample is near the image border, a line of pixels is often depressed. Lastly, the projected sample areas did not always have shapes consistent with simulation. Differences appeared in the form of protruding corners, curved edges, and varying edge length.

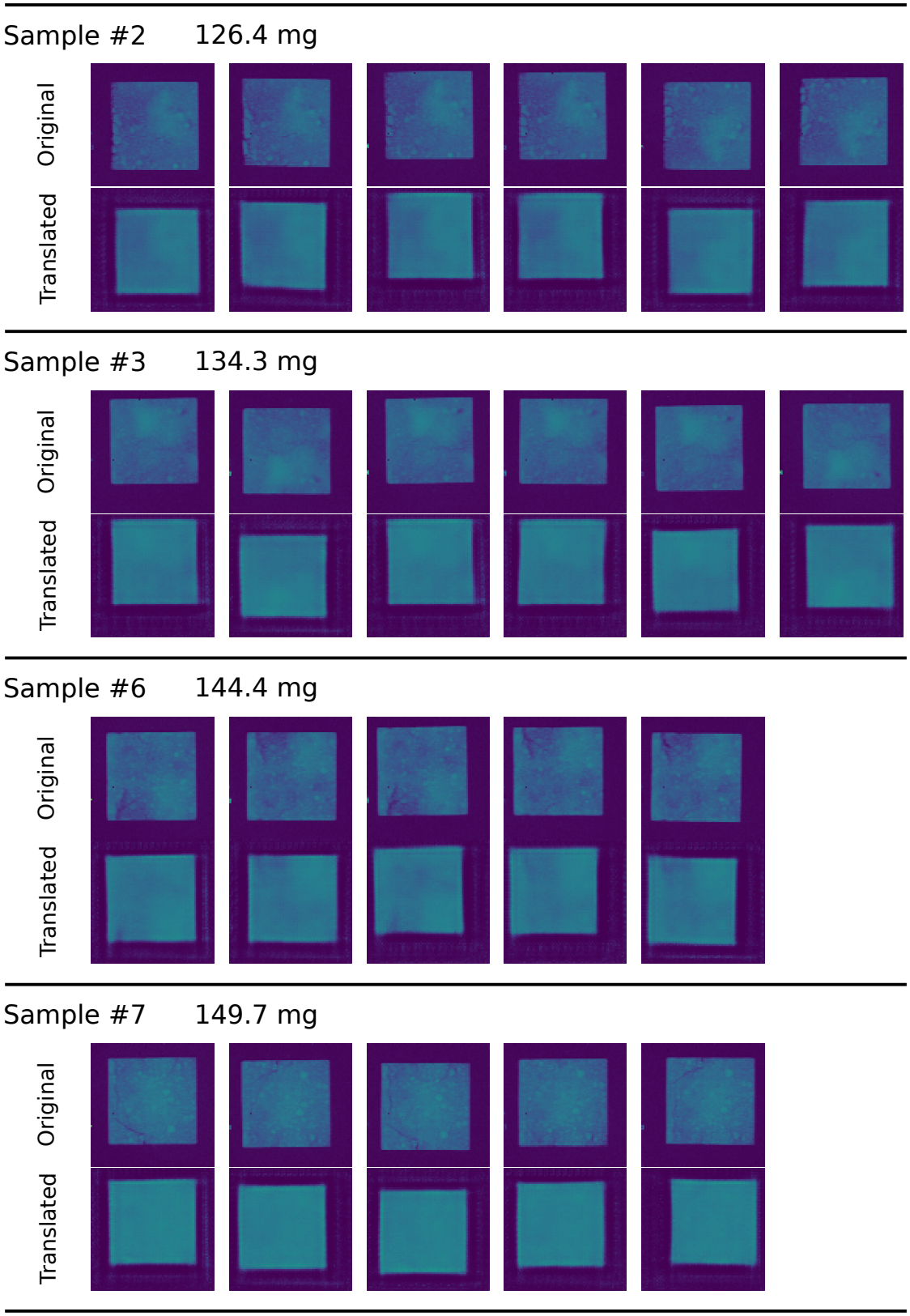

Figure 6.9: The original and translated experimental images (Samples 2, 3, 6, and 7).

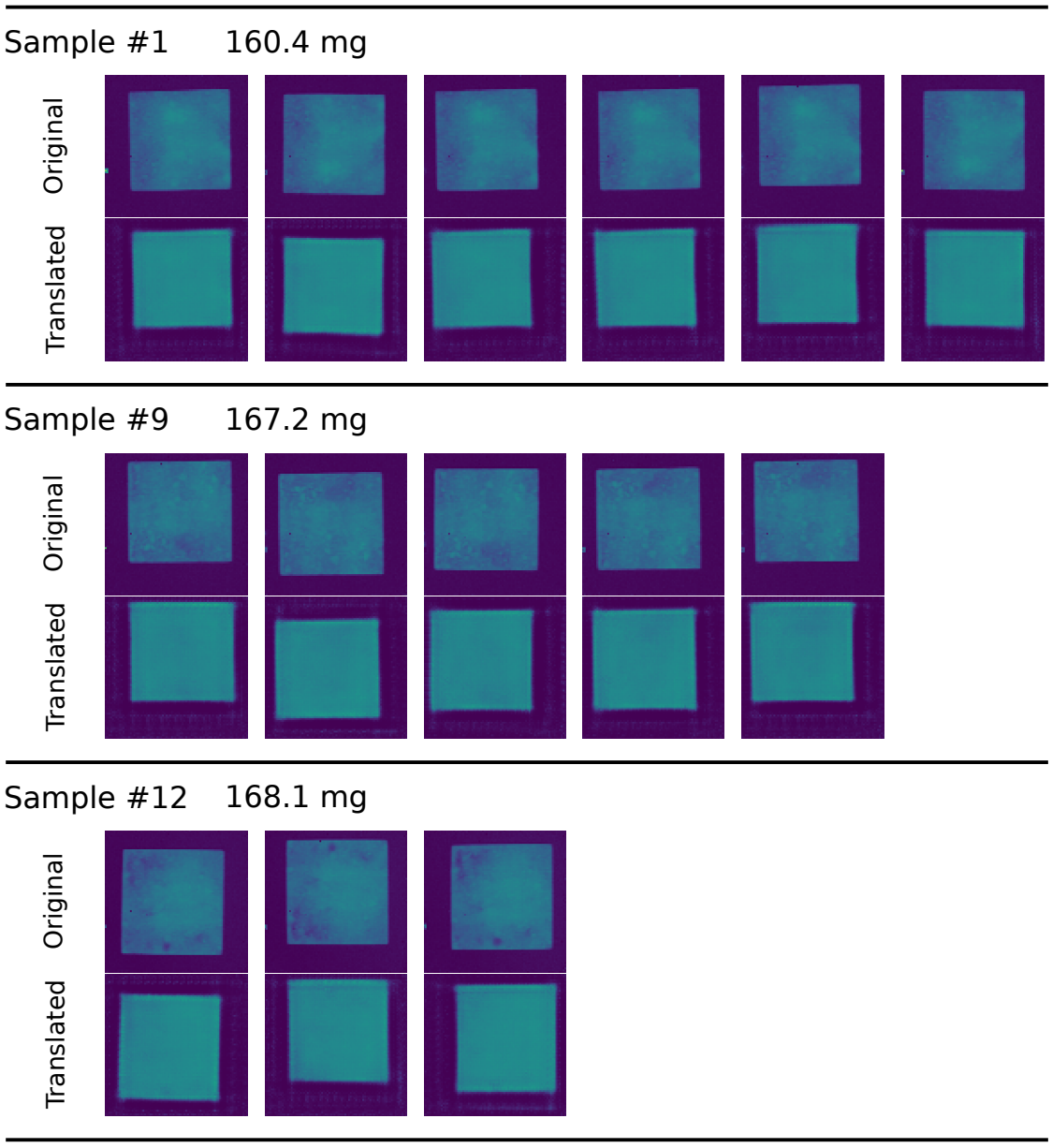

Figure 6.10: The original and translated experimental images (Samples 1, 9, and 12).

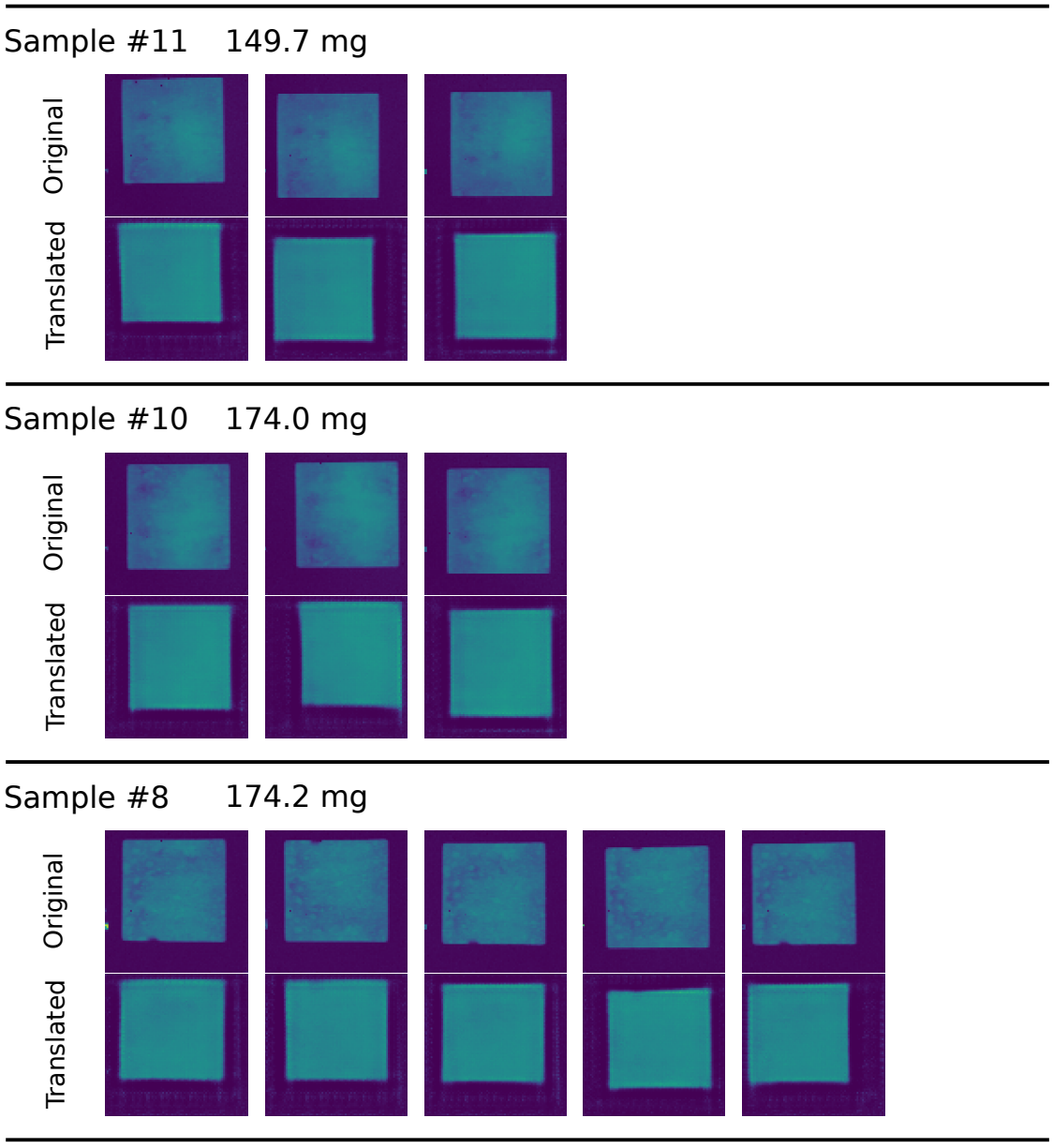

Figure 6.11: The original and translated experimental images (Samples 11, 10, and 8).

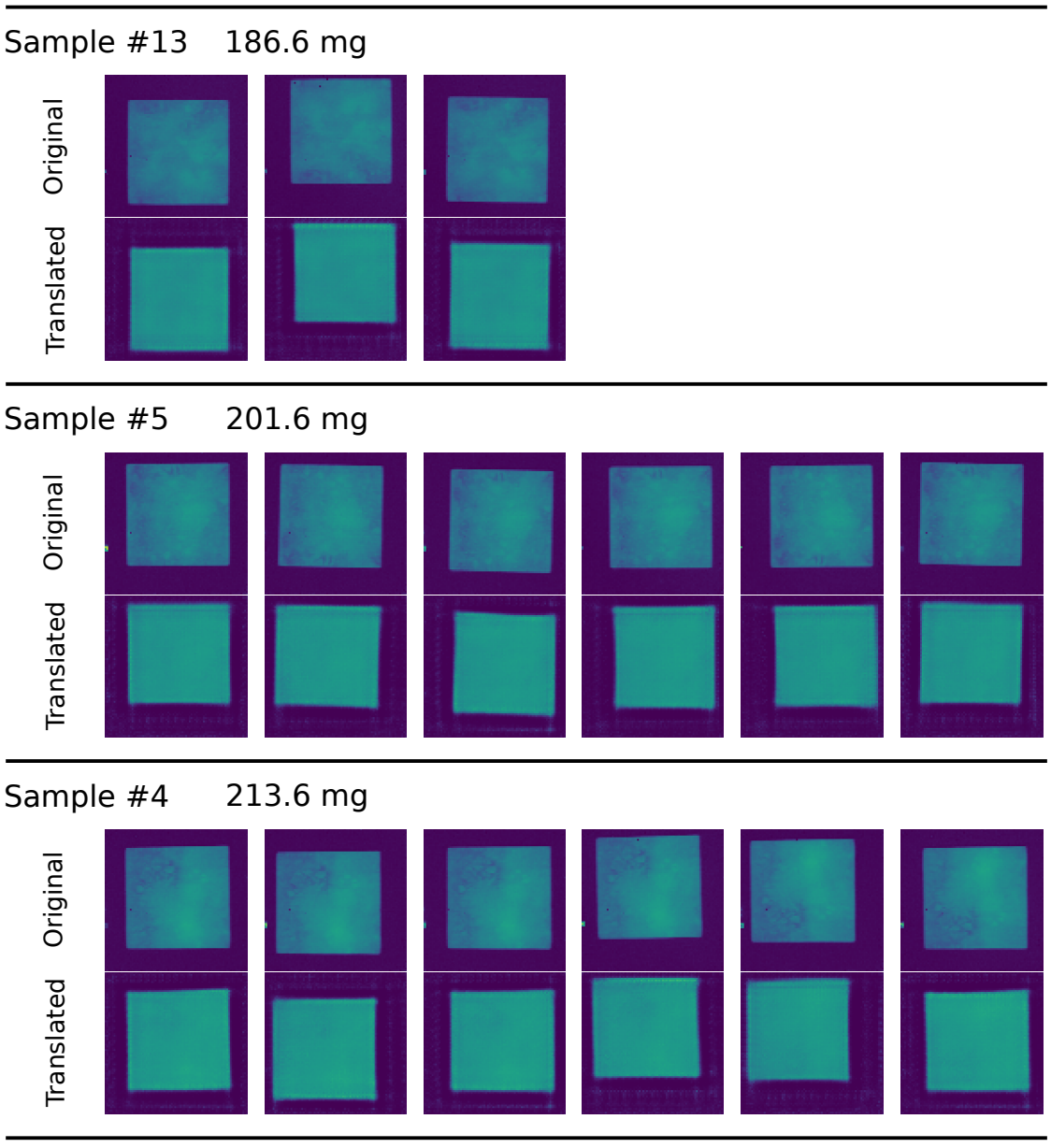

Figure 6.12: The original and translated experimental images (Samples 13, 5, and 4).

6.4.2 Supervised Regression Model

The regression model determined the elemental mass from spectral radiography data. We trained the 3D ResNet-34 architecture on a training dataset of 5000 simulated examples and a validation dataset of 1000. The training curves in Figure 6.13 show the residuals of Bi (left) and O (right) masses on the training and validation sets over the training period. The validation curves show no indication of overfitting for either element.

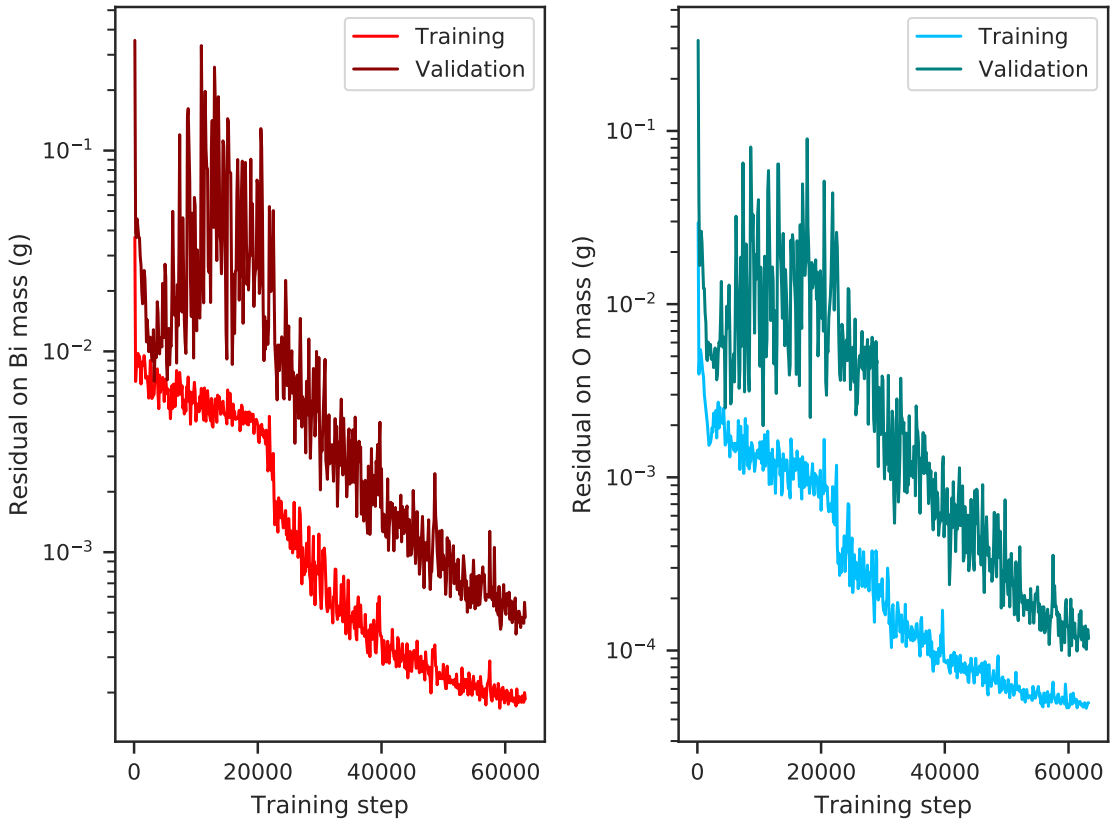

Figure 6.13: Training and validation Bi (left) and O (right) mass residuals during the training of the regression model on simulated data.

We used an additional 1000 simulations as test data to evaluate the performance of the regression model. The scatter plot in Figure 6.14 shows the total mass (the sum over elemental masses) residual vs. label. The residual had no apparent trend respect to the label value. The variance of the residual also appeared constant with residual.

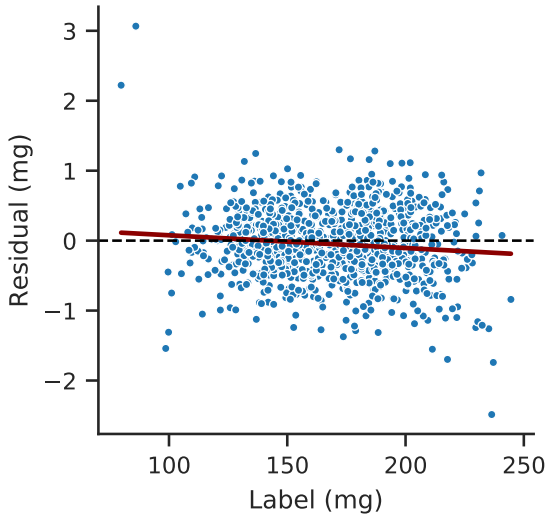

Figure 6.14: Scatter plot of residual vs. label for the regression model on simulated test data.

6.4.3 Mass Quantification of Experimental Data

The ultimate goal of this work was to quantify the mass of powder samples from experimental spectral radiographs. While the regression model quantified the elemental mass, only the total mass was available from the experimental data for verification. Therefore, we limit our analysis to the comparison of total mass by summing the calculated Bi and O mass values. The regression model was trained using only simulated data, and we tested it using three versions of the experimental data. As a baseline, we tested the raw experimental data with no domain adaptation. We also tested two versions of the domain adaptation: with and without the auxiliary regressor. The auxiliary regressor, circled with a dashed line in Figure 6.6, used the same simulation mass labels used to train the regression model and the experimental mass values from Table 6.1. While the domain adaptation training used an augmented experimental dataset, these test sets only included the original 62 experimental measurements.

The total mass residuals are plotted against the experimentally measured mass (labels) for the three test sets in Figure 6.15. The regression model consistently underpredicted mass values on raw experimental data. The residuals were negatively correlated with the

mass labels indicating that samples with higher mass were underpredicted by a larger margin. The predicted mass values on the both sets of translated data were overestimated on average. The translation without an auxiliary regressor had residuals that were more weakly correlated with mass than the raw or translated with auxiliary regressor cases. The inclusion of the auxiliary regressor both added a mass-dependent bias and increased the residual variance for predictions on measurements of the same sample.

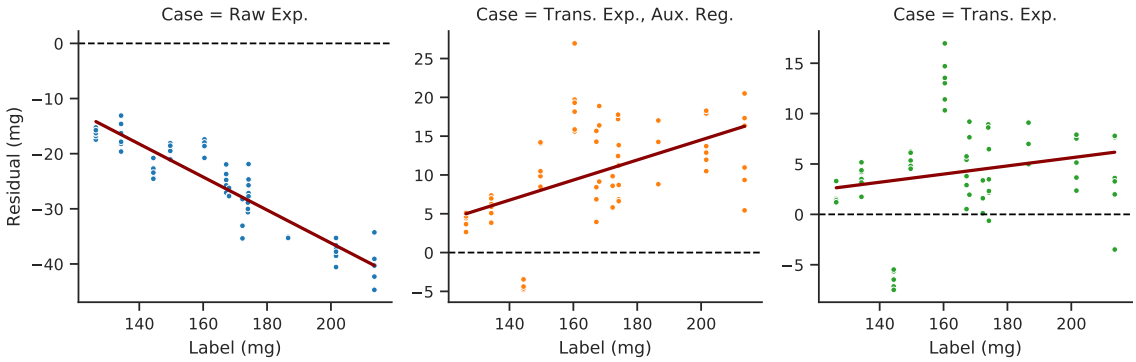

Figure 6.15: Scatter plot of residual vs. label for the regression model on experimental test data: raw experimental data (blue), translated experimental data with auxiliary regressor (orange), and translated experimental data without auxiliary regressor (green).

The three test cases in Figure 6.15 also shared some prediction behavior. Relative to the red trend lines on each plot, the bias directions of each sample were similar for each case. The relative variances of each sample were also similar. For example, the data at 160.4 g were consistently higher than the trend, whereas the data at 144.4 g were consistently below it. It was found that these variations were sensitive to the flatfield scaling described in Subsection 3.3.6.

6.5 Discussion

The stability of the domain adaptation was highly dependent on the many hyperparameters. The following had the largest impact on training outcome: the network architectures and capacities, the relative learning rates and frequencies of the two networks, the global learning rate, the use of regularization (experience replay or L1 norm), and the batch size.

There is an important balance that needs to be achieved regarding the discriminator. If the discriminator learns to recognize too quickly, the translator has no chance to learn a representation. The translator never directly sees the simulation data it is trying to mimic. Instead, it only uses the representation learned within the discriminator and is limited to learning this representation. The discriminator must therefore have a large enough capacity to provide sufficient instruction to the translator during training.

Once a stable configuration and balance was found, we experienced limitation in the translation capabilities. First, it was ultimately limited by the size and diversity of the experimental dataset. For example, the artifacts in the translated images appear to be due to averaging effects via batch normalization. The limited experimental dataset was also the motivation for a reverse domain adaptation approach. We were also limited by computing power which restricted the batch size and network depths. With more data, deep networks, and likely smaller learning rates, the translations could be sharper and more similar to simulation.

Deep, unsupervised domain adaptation has been used for tasks such as image-to-image translation that do not attempt classification or regression in addition to domain mapping. A natural question is how do we ensure the label is preserved during translation? Shrivastava *et al.* [150] addressed this with an additional L1 loss between the translated and original image to preserved the gaze direction in pictures of eyes. We found that using an L1 penalty destabilized the learning and lead to unrealistic translations. The auxiliary regressor used in this work provided means for the network to preserve label regardless of the domain. The translations, however, aligned better without the auxiliary regressor. We speculate that the regressor was in competition with the discrimination loss and worsened performance. Further, the regressor attempts to predict the same mass value for an experimental input regardless of which network is training during that step. In contrast, the discriminator output label an experimental input is different during a translator training step (1) vs. a discriminator training step (0). This difference in objective likely played a role in making

the discriminator less effective.

The procedure for scaling the flatfield likely introduced additional variance to the data that was insurmountable for this approach. We were motivated by a time constraint on data acquisition to take a limited number of flatfield measurements. Changes to the system geometry or to the acquisition parameters across measurement days can cause violations to the assumptions in Subsection 3.3.6. Ideally, there would be a flatfield measurement taken immediately before or after a sample measurement. Alternatively, a flatfield for each day of measurements would mitigate any changes to the system. More frequent flatfield measurements would come at the price of fewer sample measurements, which is the trade-off we faced.

CHAPTER 7
CONCLUSION

This thesis explored elemental mass quantification using spectral X-ray radiography data. The recent advances in pixelated, spectroscopic X-ray detectors are now able to provide spatio-spectral information that was not available before. A growing interest has revealed numerous applications of spectral X-ray radiography in safeguards, national security, industry, medicine, and research. We focused on quantitative methods for solid samples containing high-Z materials.

In Chapter 4, we examined an existing method that uses nonlinear least squares to quantify elemental mass from spectral X-ray radiography data. The physics model used in this method assumed uniform exponential attenuation across each pixel face. We showed that this assumption leads to mass quantification errors, particularly at the edges of the sample where the pixel only partially views the sample. To address this problem, we updated the physics model to allow for partial-volume attenuation. We updated the algorithm to also solve for the new fraction parameter and presented additional stabilization techniques that were necessary for the partial-fraction physics model. We found that the updated physics model made the problem more ill-posed, and simple constraints to the gradient and fraction were sufficient to stabilize the optimization process. We demonstrated improved mass quantification accuracy on simulations of UO_2 samples with the updated physics model and algorithm.

In Chapter 5, we presented a deep learning alternative to the traditional algorithmic approach. A successful deep learning model has the potential to improve the operational viability of a spectral X-ray radiography system. By including variations in system geometry or statistical noise, we hypothesized that the model would accurately predict elemental mass on data subject to these variations. We also hypothesized that a deep learning model

would facilitate the fusion of spectral X-ray radiography data with X-ray fluorescence data. These hypotheses were tested using synthesized datasets that incorporated systematic variations and statistical noise. The deep learning models on simulated spectral X-ray radiography performed with high accuracy. The models trained on only spectral X-ray radiography data or XRF data degraded in performance with increasing number of elements to quantify. The multimodal model, however, showed better performance than its two input modes.

The deep learning approach offers advantages over the traditional algorithmic approach. First, by including geometric variation in the training data, we were able to train a model that did not require high-confidence knowledge of the geometry. This is an attractive quality for the deployment of such a system to industry. Second, the ability to incorporate XRF data is beneficial for the identification or quantification of trace elements that would be outside the sensitivity of the transmission system and for tasks that must quantify multiple elements. Third, we found that the algorithmic approach experienced performance loss in the presence of statistical noise, but the deep learning models were all trained on noise equivalent to 10 min acquisitions and had strong performance. This suggests the use of a deep learning model would permit shorter acquisitions and, therefore, would have higher throughput.

A major drawback of the deep learning approach is the creation of training data. Using simulated data is cheaper and faster than collecting experimental measurements, and the simulations are labeled (with the elemental masses). There exists a domain shift, however, between simulation and experiment that must be addressed. The algorithmic approach has the advantage of not being limited to the distribution spanned by training data. Instead, a well-characterized system geometry, source, and detector combined with the radiation transport physics model provides a reliable way to quantify elemental mass.

This work paved the way for several research avenues. We would like to explore the extension of the method to more elements. While other techniques with the same objective become limited by number of elements to quantify, deep spectral X-ray radiography has

shown the potential to exceed the current standards. Such an implementation would require a system with a source whose energy distribution properly spans the absorption edges of interest. Additionally, we are interested in exploring the use of a pixelated, spectroscopic X-ray detector or multiple detectors for the XRF measurements. The addition of XRF improved the overall performance; it is likely there is useful information in the signal that has yet to be exploited.

In Chapter 6, we extended the deep learning approach to experimental data with the addition of deep unsupervised domain adaptation. This adversarial approach learned to map from the experimental domain to the simulation domain using unpaired simulations and experimental measurements of Bi_2O_3 powder. The experimental measurements were then translated such that they were indistinguishable from simulated data by a discriminator neural network. A regression network similar to those used in Chapter 5 was trained on simulated data for mass quantification. We then quantified the mass of the translated experimental data using the regression network which were compared to the experimentally measured masses. The relative error on the translated experimental data ($< 10\%$) was about an order of magnitude worse than on the simulated test data ($< 1\%$).

The transition to experimental data poses a challenge for both deep learning and algorithmic methods alike. While the deep learning model can be trained to learn variations present in the experimental system, they must be known and incorporated into the training dataset. Otherwise, the model will not properly handle them. This was the case regarding the flatfield data used in Chapter 6 which limited the performance. Further, the limited number and diversity of experimental data restricted the capacity of the translation stage. More data would permit mapping to and from both domains, but we were limited to mapping experimental data to the simulation domain.

In future experiments, we are interested in taking additional precautions to ensure high quality experimental data. First, flatfield measurements must accompany all measurements of samples. This would help address the fluctuations that occur day-to-day as well as within

the same day. Second, both deep learning-based and physics-based methods would benefit from the inclusion of reference samples for calibration. The references could be taken either as separate measurements or taken simultaneously with a sample.

Appendices

APPENDIX A

PROOF OF UNDERESTIMATION OF AREAL DENSITY

A single pixel views a material with mass attenuation coefficient $\mu \in \mathbb{R}_{\geq 0}$. A fraction of the projected area α is attenuated by areal density ρ_1 and the remaining fraction $(1-\alpha)$ is attenuated by the same material of a different areal density ρ_2, where $\alpha \in [0,1]$ and all densities $\rho \in \mathbb{R}_{\geq 0}$. The true, total areal density ρ_{true} used to determine the mass is

$$\rho_{\text{true}} = \alpha\rho_1 + (1-\alpha)\rho_2. \tag{A.1}$$

The total flux Φ_{obs} observed by the pixel is the linear combination of the attenuation from each areal density, where $\Phi \in \mathbb{R}_{\geq 0}$. This is given by

$$\Phi_{\text{obs}} = \alpha\Phi_0 \exp(-\mu\rho_1) + (1-\alpha)\Phi_0 \exp(-\mu\rho_2), \tag{A.2}$$

where Φ_0 is the unattenuated flux. The algorithm attempts to find some effective areal density ρ_{eff} by setting the constraint that

$$\Phi_{\text{obs}} = \Phi_0 \exp(-\mu\rho_{\text{eff}}). \tag{A.3}$$

Definition A.1 *A function $f : \mathbb{R} \mapsto \mathbb{R}$ is convex if its domain X is a convex set and $\forall x_1, x_2 \in X$, $\alpha \in [0,1]$,*

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2) \tag{A.4}$$

Definition A.2 *A twice differentiable function is convex on an interval iff its second derivative is nonnegative on that interval.*

Lemma A.1 *For some ρ_{eff} used to estimate Φ_{obs} as (Equation A.3) where Φ_{obs} is given by*

(Equation A.2), the true areal density ρ_{true} given by (Equation A.1) is greater than or equal to ρ_{eff}.

Proof The twice differentiable function $\Phi(\rho) = \Phi_0 \exp(-\mu\rho)$ with $\Phi_0 \geq 0$ has second derivative

$$\frac{d^2\Phi}{d\rho^2} = \mu^2 \Phi_0 \exp(-\mu\rho). \tag{A.5}$$

For all $\rho \in \mathbb{R}$, $\frac{d^2\Phi}{d\rho^2} \geq 0$; therefore, $\Phi(\rho)$ is convex. From the first definition of convexity,

$$\begin{aligned}
\Phi(\alpha\rho_1 + (1-\alpha)\rho_2) &\leq \alpha\Phi(\rho_1) + (1-\alpha)\Phi(\rho_2) \\
\Phi_0 \exp[-\mu(\alpha\rho_1 + (1-\alpha)\rho_2)] &\leq \alpha\Phi_0 \exp(-\mu\rho_1) + (1-\alpha)\Phi_0 \exp(-\mu\rho_2) \\
\Phi_0 \exp(-\mu\rho_{\text{true}}) &\leq \Phi_0 \exp(-\mu\rho_{\text{eff}}) \\
\rho_{\text{true}} &\geq \rho_{\text{eff}}.
\end{aligned} \tag{A.6}$$

Therefore, for all ρ_1 and ρ_2, the true areal density is greater than or equal to the effective areal density. This indicates that an algorithm approximating attenuation by a single effective areal density will underestimate the true areal density.

APPENDIX B

HYPERPARAMETER GRID SEARCH

The relative error on total uranium and total oxygen mass is presented here with a confidence of one standard deviation.

Table B.1: Performance Without Stabilization

f	ε	m_U rel. error	m_O rel. error	# iter.
εx	10^0	–	–	–
	10^1	–	–	–
	10^2	$.069\% \pm .050\%$	$4.2\% \pm 2.6\%$	12 ± 0.0
εx^2	10^0	$.097\% \pm .053\%$	$4.1\% \pm 2.6\%$	$9.9 \pm .74$
	10^1	$.067\% \pm .050\%$	$4.1\% \pm 2.6\%$	8.5 ± 1.6
	10^2	$.068\% \pm .050\%$	$4.1\% \pm 2.6\%$	$9.2 \pm .83$
εx^3	10^0	–	–	–
	10^1	$.067\% \pm .050\%$	$4.1\% \pm 2.6\%$	$7.5 \pm .50$
	10^2	$.065\% \pm .050\%$	$4.2\% \pm 2.6\%$	$7.2 \pm .56$
$e^{\varepsilon x}$	10^0	–	–	–
	10^1	$.064\% \pm .050\%$	$4.2\% \pm 2.6\%$	$8.0 \pm .20$
	10^2	$.065\% \pm .050\%$	$4.1\% \pm 2.6\%$	12 ± 4.1
$\sigma(\varepsilon x)$	10^0	–	–	–
	10^1	–	–	–
	10^2	$.065\% \pm .052\%$	$4.2\% \pm 2.6\%$	11 ± 1.1

Table B.2: Bounded Fraction Performance

f	ε	f_{min}	m_U rel. error	m_O rel. error	# iter.
εx	10^0	10^{-3}	$.12\% \pm .053\%$	$4.0\% \pm 2.6\%$	13 ± 1.1
	10^1	10^{-5}	$.068\% \pm .050\%$	$4.2\% \pm 2.6\%$	$8.2 \pm .81$
	10^2	10^{-5}	$.069\% \pm .050\%$	$4.2\% \pm 2.6\%$	12 ± 0.0
εx^2	10^0	10^{-5}	$.097\% \pm .053\%$	$4.1\% \pm 2.6\%$	$8.7 \pm .46$
	10^1	10^{-5}	$.069\% \pm .050\%$	$4.1\% \pm 2.6\%$	$6.2 \pm .36$
	10^2	10^{-5}	$.069\% \pm .050\%$	$4.1\% \pm 2.6\%$	12 ± 6.7
εx^3	10^0	10^{-5}	$.086\% \pm .052\%$	$4.1\% \pm 2.6\%$	$7.6 \pm .49$
	10^1	10^{-5}	$.070\% \pm .050\%$	$4.1\% \pm 2.6\%$	$6.2 \pm .64$
	10^2	10^{-5}	$.069\% \pm .050\%$	$4.1\% \pm 2.6\%$	$6.2 \pm .43$
$e^{\varepsilon x}$	10^0	10^{-4}	$.096\% \pm .053\%$	$4.1\% \pm 2.6\%$	13 ± 2.2
	10^1	10^{-5}	$.064\% \pm .050\%$	$4.2\% \pm 2.6\%$	$8.0 \pm .20$
	10^2	10^{-5}	$.064\% \pm .051\%$	$4.2\% \pm 2.6\%$	12 ± 4.2
$\sigma(\varepsilon x)$	10^0	10^{-3}	–	–	–
	10^1	10^{-5}	–	–	–
	10^2	10^{-5}	$.066\% \pm .052\%$	$4.2\% \pm 2.6\%$	11 ± 1.0

Table B.3: Gradient Threshold Performance

f	ε	g_{th}	m_U rel. error	m_O rel. error	# iter.
εx	10^0	10^{-9}	$.12\% \pm .053\%$	$4.0\% \pm 2.6\%$	$10. \pm .22$
	10^1	10^{-9}	$.069\% \pm .050\%$	$4.1\% \pm 2.6\%$	$8.2 \pm .82$
	10^2	10^{-9}	$.069\% \pm .050\%$	$4.1\% \pm 2.6\%$	12 ± 0.0
εx^2	10^0	10^{-9}	$.097\% \pm .053\%$	$4.5\% \pm 2.7\%$	11 ± 1.2
	10^1	10^{-9}	$.067\% \pm .050\%$	$4.2\% \pm 2.6\%$	8.4 ± 1.2
	10^2	10^{-9}	$.068\% \pm .050\%$	$4.2\% \pm 2.6\%$	$9.7 \pm .90$
εx^3	10^0	10^{-9}	$.084\% \pm .051\%$	$5.6\% \pm 3.5\%$	$10. \pm 2.0$
	10^1	10^{-9}	$.065\% \pm .050\%$	$4.3\% \pm 2.6\%$	$7.9 \pm .50$
	10^2	10^{-9}	$.066\% \pm .050\%$	$4.2\% \pm 2.6\%$	$8.2 \pm .63$
$e^{\varepsilon x}$	10^0	10^{-9}	$.087\% \pm .051\%$	$15\% \pm 7.5\%$	9.5 ± 2.3
	10^1	10^{-9}	$.068\% \pm .050\%$	$4.2\% \pm 2.6\%$	$11 \pm .71$
	10^2	10^{-9}	$.057\% \pm .050\%$	$4.3\% \pm 2.6\%$	12 ± 3.4
$\sigma(\varepsilon x)$	10^0	10^{-9}	$.077\% \pm .059\%$	$4.3\% \pm 3.1\%$	34 ± 2.1
	10^1	10^{-9}	$.046\% \pm .14\%$	$9.1\% \pm 5.1\%$	37 ± 5.3
	10^2	10^{-9}	$.057\% \pm .053\%$	$4.3\% \pm 2.6\%$	$12 \pm .87$

APPENDIX C

SIMULATION GEOMETRY AND DENSITY SAMPLING DETAILS

100 μm $\times$ 100 μm $\times$ 1 mm voxels

$\Delta x \sim \mathcal{U}(-2 \text{ mm}, 2 \text{ mm})$, $\Delta y \sim \mathcal{U}(-2 \text{ mm}, 2 \text{ mm})$, $\Delta z \sim \mathcal{U}(-0.5 \text{ mm}, 0.5 \text{ mm})$, $\Delta r_1 \sim \mathcal{U}(-0.5 \text{ rad}, 0.5 \text{ rad})$, $\Delta r_2 \sim \mathcal{U}(-1 \text{ rad}, 1 \text{ rad})$

Algorithm 1 Sampling total voxel densities for a simulation

procedure CALCTOTALDENSITY(d_1, d_2, f)
    $\rho_t \leftarrow \frac{d_1 d_2}{d_1 f + d_2(1-f)}$
    return ρ_t

Algorithm 2 Sampling total voxel densities for a simulation

procedure SAMPLEBI
    sample $B \sim \mathcal{N}\left(1.5\frac{g}{cm^3}, (0.75\frac{g}{cm^3})^2\right)$
    sample $\sigma_B^2 \sim \mathcal{U}\left((0.015\frac{g}{cm^3})^2, (0.75\frac{g}{cm^3})^2\right)$
    for voxel j **do**
        sample $b_j \sim \mathcal{N}\left(B, \sigma_B^2\right)$
        $f_j \leftarrow 0$
        $\rho_j \leftarrow$ CALCTOTALDENSITY(b_j, 1, f_j)
    return ρ, f

Algorithm 3 Sampling total voxel densities for a simulation

procedure SAMPLEBIGD
    sample $B \sim \mathcal{N}\left(1.5\frac{g}{cm^3}, (0.75\frac{g}{cm^3})^2\right)$
    sample $\sigma_B^2 \sim \mathcal{U}\left((0.015\frac{g}{cm^3})^2, (0.75\frac{g}{cm^3})^2\right)$
    sample $G \sim \mathcal{N}\left(1.05\frac{g}{cm^3}, (0.25\frac{g}{cm^3})^2\right)$
    sample $F \sim \mathcal{U}(0, 0.1)$
    $\sigma_F^2 \leftarrow (0.1F)^2$
    for voxel j **do**
        sample $b_j \sim \mathcal{N}\left(B, \sigma_B^2\right)$
        sample $f_j \sim \mathcal{N}\left(F, \sigma_F^2\right)$
        $\rho_j \leftarrow$ CALCTOTALDENSITY(b_j, G, f_j)
    return ρ, f

Algorithm 4 Sampling total voxel densities for a simulation

procedure SAMPLEBIW
    sample $B \leftarrow 1.5\frac{g}{cm^3}$
    sample $W \sim \mathcal{U}(1.4\frac{g}{cm^3}, 6.5\frac{g}{cm^3})$
    sample $\sigma_W^2 \sim \mathcal{U}\left((0.01W)^2, (0.5W)^2\right)$
    sample $F \sim \mathcal{U}(0, 1)$
    for voxel j **do**
        sample $w_j \sim \mathcal{N}\left(W, \sigma_W^2\right)$
        $f_j \leftarrow F$
        $\rho_j \leftarrow$ CALCTOTALDENSITY(B, w_j, f_j)
    return ρ, f

Algorithm 5 Sampling total voxel densities for a simulation

procedure GENERATESIMULATIONS
    $n \leftarrow 10000$ ▷ Num. simulations
    for $i \leftarrow 1, n$ **do**
        $\rho_B, f_B \leftarrow$ SAMPLEBI
        SIMULATE(Bi, ρ_B, f_B)
        $\rho_G, f_G \leftarrow$ SAMPLEBIGD
        SIMULATE(Bi + Gd, ρ_G, f_G)
        $\rho_W, f_W \leftarrow$ SAMPLEBIW
        SIMULATE(Bi + W, ρ_W, f_W)

REFERENCES

[1] "The Treaty on the Non-Proliferation of Nuclear Weapons," 1968.

[2] *Safeguards Techniques and Equipment*. No. 1 (Rev. 2) in International Nuclear Verification Series, Vienna, Austria: International Atomic Energy Agency, 2011.

[3] P. A. Russo, S. T. Hsue, J. K. Sprinkle, S. S. Johnson, and Y. Asakura, "Transmissions for Precise K-Edge and Passive Assay of Plutonium Concentration and Isotopic Fractions in Product Solutions," Report LA–9440-MS, Los Alamos National Lab., United States, Aug. 1982.

[4] T. F. Guzzardo, R. D. McElroy, S. Croft, J. Garrison, R. Venkataraman, and C. A. Pickett, "Stability of Working Reference Standards for Hybrid K-Edge Densitometer Quality Assurance," in *IAEA Safeguards Symposium*, vol. 4, (Vienna, Austria), International Atomic Energy Agency, 8 2014.

[5] A. Berlizov, A. Schachinger, K. Roetsch, N. Erdmann, H. Schorl, M. Vargas, J. Zsigrai, A. Kulko, M. Keselica, F. Caillou, V. Unsal, and A. Walczak-Typke, "Feedback from operational experience of on-site deployment of bias defect analysis with COMPUCEA," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 307, pp. 1901–1909, Mar. 2016.

[6] "In situ applications of x ray fluorescence techniques: final report of a coordinated research project, 2000-2003," Tech. Rep. IAEA-TECDOC-1456, International Atomic Energy Agency, Vienna, Austria, 2005.

[7] A. J. Gilbert, B. S. McDonald, S. M. Robinson, K. D. Jarman, T. A. White, and M. R. Deinert, "Non-invasive material discrimination using spectral x-ray radiography," *Journal of Applied Physics*, vol. 115, p. 154901, Apr. 2014.

[8] Y. Wang, C. Lin, and J. Miller, "Quantitative analysis of exposed grain surface area for multiphase particles using X-ray microtomography," *Powder Technology*, vol. 308, pp. 368–377, Feb. 2017.

[9] M. Soret, S. L. Bacharach, and I. Buvat, "Partial-Volume Effect in PET Tumor Imaging," *Journal of Nuclear Medicine*, vol. 48, pp. 932–945, June 2007.

[10] X. Liu, L. Yu, A. N. Primak, and C. H. McCollough, "Quantitative imaging of element composition and mass fraction using dual-energy CT: Three-material decomposition: Three-material decomposition of mass fraction," *Medical Physics*, vol. 36, pp. 1602–1609, Apr. 2009.

[11] T. Sellerer, S. Ehn, K. Mechlem, M. Duda, M. Epple, P. B. Nol, and F. Pfeiffer, "Quantitative dual-energy micro-CT with a photon-counting detector for material science and non-destructive testing," *PLOS ONE*, vol. 14, p. e0219659, July 2019.

[12] C. K. Egan, S. D. M. Jacques, M. D. Wilson, M. C. Veale, P. Seller, A. M. Beale, R. A. D. Pattrick, P. J. Withers, and R. J. Cernik, "3D chemical imaging in the laboratory by hyperspectral X-ray computed tomography," *Scientific Reports*, vol. 5, Dec. 2015.

[13] F. H. Attix, *Introduction to Radiological Physics and Radiation Dosimetry*. New York: Wiley, 1986.

[14] M. J. Berger, J. H. Hubbell, S. M. Seltzer, J. Chang, J. S. Coursey, R. Sukumar, D. S. Zucker, and K. Olsen, *XCOM: Photon Cross Section Database (version 1.5)*. National Institute of Standards and Technology, 2010.

[15] M. A. Laskey and D. Phil, "Dual-Energy X-Ray Absorptiometry and Body Composition," *Nutrition*, vol. 12, p. 7, Jan. 1996.

[16] G. Deng, M. Chen, P. He, X. Wang, X. Wu, X. Guo, P. Li, B. Wei, K. An, X. Zheng, and P. Feng, "The Experimental Study on Geometric Calibration and Material Discrimination for In Vivo Dual-Energy CT Imaging," *BioMed Research International*, vol. 2019, pp. 1–8, May 2019.

[17] A. Panahifar, N. Samadi, T. M. Swanston, L. D. Chapman, and D. M. Cooper, "Spectral K-edge subtraction imaging of experimental non-radioactive barium uptake in bone," *Physica Medica*, vol. 32, pp. 1765–1770, Dec. 2016.

[18] A. J. Gilbert, B. S. McDonald, and L. E. Smith, "Spectral X-ray Radiography for Safeguards at Nuclear Fuel Fabrication Facilities: A Feasibility Study," in *International Conference on Mathematics & Computational Methods Applied to Nuclear Science & Engineering*, (Jeju, Korea), Apr. 2017.

[19] N. Ducros, J. F. P.-J. Abascal, B. Sixou, S. Rit, and F. Peyrin, "Regularization of nonlinear decomposition of spectral x-ray projection images," *Medical Physics*, vol. 44, pp. e174–e187, Sept. 2017.

[20] V. Rebuffel, G. Beldjoudi, A. Brambilla, and J. Rinkel, "Material Identification Methods in Spectral Radiography Using CdTe Semiconductor Detectors," p. 9, June 2011.

[21] A. Brambilla, A. Gorecki, A. Potop, C. Paulus, and L. Verger, "Basis material decomposition method for material discrimination with a new spectrometric X-ray imaging detector," *Journal of Instrumentation*, vol. 12, pp. P08014–P08014, Aug. 2017.

[22] A. Raja, M. Moghiseh, C. Bateman, N. de Ruiter, B. Schon, N. Schleich, T. Woodfield, A. Butler, and N. Anderson, "Measuring Identification and Quantification Errors in Spectral CT Material Decomposition," *Applied Sciences*, vol. 8, p. 467, Mar. 2018.

[23] X. Wu, Q. Wang, J. Ma, W. Zhang, P. Li, and Z. Fang, "A hyperspectral X-ray computed tomography system for enhanced material identification," *Review of Scientific Instruments*, vol. 88, p. 083111, Aug. 2017.

[24] E. S. Jimenez, K. R. Thompson, R. N. Goodner, and A. Stohn, "Leveraging multichannel x-ray detector technology to improve quality metrics for industrial and security applications," p. 15, SPIE, Sept. 2017.

[25] D. Wu, L. Zhang, X. Zhu, X. Xu, and S. Wang, "A weighted polynomial based material decomposition method for spectral x-ray CT imaging," *Physics in Medicine and Biology*, vol. 61, pp. 3749–3783, May 2016.

[26] C. K. Egan, M. D. Wilson, M. C. Veale, P. Seller, S. D. Jacques, and R. J. Cernik, "Material specific X-ray imaging using an energy-dispersive pixel detector," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 324, pp. 25–28, Apr. 2014.

[27] X. Wang, D. Meier, K. Taguchi, D. J. Wagenaar, B. E. Patt, and E. C. Frey, "Material separation in x-ray CT with energy resolved photon-counting detectors: Material separation with energy resolved photon-counting detectors," *Medical Physics*, vol. 38, pp. 1534–1546, Feb. 2011.

[28] K. Jones, B. Gordon, A. Hanson, J. Hastings, M. Howells, H. Kraner, and J. Chen, "Application of synchrotron radiation to elemental analysis," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 3, pp. 225–231, Apr. 1984.

[29] M. Cotte, J. Susini, J. Dik, and K. Janssens, "Synchrotron-Based X-ray Absorption Spectroscopy for Art Conservation: Looking Back and Looking Forward," *Accounts of Chemical Research*, vol. 43, pp. 705–714, June 2010.

[30] V. Tichy, T. Holy, J. Jakubek, V. Linhart, S. Pospisil, and Z. Vykydal, "X-ray fluorescence imaging with pixel detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 591, pp. 67–70, June 2008.

[31] E. Liotti, A. Lui, T. Connolley, I. Dolbnya, K. Sawhney, A. Malandain, M. Wilson, M. Veale, P. Seller, and P. Grant, "Mapping of multi-elements during melting and solidification using synchrotron X-rays and pixel-based spectroscopy," *Scientific Reports*, vol. 5, Dec. 2015.

[32] M. D. Wilson, T. Connolley, I. P. Dolbnya, P. S. Grant, E. Liotti, A. Lui, A. Malandain, K. Sawhney, P. Seller, and M. C. Veale, "Energy dispersive detector for white beam synchrotron x-ray fluorescence imaging," *AIP Conference Proceedings*, vol. 1741, no. 1, p. 050008, 2016.

[33] J. Py, J.-E. Groetz, J.-C. Hubinois, and D. Cardona, "Determination of plutonium in nitric acid solutions using energy dispersive L X-ray fluorescence with a low

power X-ray generator," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 780, pp. 131–137, Apr. 2015.

[34] H. Ottmar and H. Eberle, "The hybrid K-edge/K-XRF densitometer: principles - design - performance," Report KFK–4590, Kernforschungszentrum Karlsruhe GmbH, Germany, Feb. 1991.

[35] C. G. Worley and G. J. Havrilla, "Micro-X-ray Fluorescence Characterization of Mixed Oxide Fuel Surrogate Feed Material," *Analytical Chemistry*, vol. 70, pp. 2957–2963, July 1998.

[36] C. G. Worley, G. J. Havrilla, and P. S. Dunn, "Quantification of Large Scale Micro-X-Ray Fluorescence Elemental Images," *Applied Spectroscopy*, vol. 55, pp. 1448–1454, Nov. 2001.

[37] V. Natarajan, B. Rajeswari, B. A. Dhawale, N. S. Hon, S. V. Godbole, and V. K. Manchanda, "Application of energy dispersive X-ray fluorescence for the determination of metallic impurities in ThO2," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 280, pp. 27–31, Apr. 2009.

[38] C. G. Worley, "Analysis of nuclear materials by energy dispersive X-ray fluorescence and spectral effects of alpha decay," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 282, pp. 539–542, Nov. 2009.

[39] S. Biswas, V. H. Rupawate, K. N. Hareendran, and S. B. Roy, "Determination of iron in uranium matrix using energy dispersive X-ray fluorescence (EDXRF) technique," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 306, pp. 543–548, Nov. 2015.

[40] J. M. Maia, R. M. Curado da Silva, and Y.-S. Kim, "Prospects on Low-Z Elements K Fluorescence and Actinide-Radionuclides L Fluorescence X-Ray Detection With Cooled CZT," *IEEE Transactions on Nuclear Science*, vol. 62, pp. 577–587, Apr. 2015.

[41] A. Pandey, S. Dhara, F. A. Khan, A. Kelkar, P. Kumar, R. B. Bhatt, and P. G. Behere, "Analysis of Th and U in thorium-based mixed-oxide fuel using wavelength dispersive X-ray fluorescence spectrometer," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 319, pp. 775–781, Mar. 2019.

[42] A. Pandey, S. Dhara, F. A. Khan, A. Kelkar, R. Yadav, P. Kumar, D. B. Sathe, R. B. Bhatt, and P. G. Behere, "Direct determination of uranium in sintered deeply depleted uranium oxide pellets by wavelength dispersive X-ray fluorescence spectrometry," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 323, pp. 275–281, Jan. 2020.

[43] "Design, Development and Optimization of a Low Cost System for Digital Industrial Radiology," Tech. Rep. 2, IAEA, Vienna, Austria, 2013.

[44] M. Wasim, A. Tariq, M. A. Shafique, and R. N. Qureshi, "Characterization and differentiation of iron ores using X-ray diffractometry, k0 instrumental neutron activation analysis and inductively coupled plasma optical emission spectrometry," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 323, pp. 179–187, Jan. 2020.

[45] S. Pathak, P. Das, A. K. Sikdar, J. Nandi, S. Bhattacharyya, T. Bhattacharjee, S. Bhattacharya, S. S. Alam, and A. Ray, "Single crystal HPGe (80%) versus BGO shielded CLOVER detector for high precision decay rate measurements: a comparative study," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 323, pp. 1353–1358, Mar. 2020.

[46] A. Leskinen, S. Salminen-Paatero, A. Rty, M. Tanhua-Tyrkk, T. Iso-Markku, and E. Puukko, "Determination of 14C, 55Fe, 63Ni and gamma emitters in activated RPV steel samples: a comparison between calculations and experimental analysis," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 323, pp. 399–413, Jan. 2020.

[47] G. F. Knoll, *Radiation Detection and Measurement*. Wiley, 4th ed., 8 2010.

[48] N. Tsoulfanidis, *Measurement and Detection of Radiation*. CRC Press, 3rd ed., 12 2010.

[49] D. J. Griffiths, *Introduction to Quantum Mechanics*. Cambridge University Press, 2nd ed., 8 2016.

[50] M. Niraula, A. Nakamura, T. Aoki, Y. Tomita, and Y. Hatanaka, "Stability issues of high-energy resolution diode type cdte nuclear radiation detectors in a long-term operation," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 491, no. 1, pp. 168 – 175, 2002.

[51] L. Verger, J. P. Bonnefoy, F. Glasser, and P. Ouvrier-Buffet, "New developments in cdte and cdznte detectors for x and -ray applications," *Journal of Electronic Materials*, vol. 26, pp. 738–744, 1997.

[52] M. D. Wilson, P. Barnes, R. C. Cernik, C. C. T. Hansson, S. Jacques, L. L. Jones, P. Seller, P. J. Sellin, T. Sochi, M. C. Veale, P. Veeramani, P. J. Withers, and C. P. Youd, "Comparison of the X-ray performance of small pixel CdTe and CZT detectors," pp. 3942–3946, IEEE, Oct. 2010.

[53] P. Siffert, J. Berger, C. Scharager, A. Cornet, R. Stuck, R. O. Bell, H. B. Serreze, and F. V. Wald, "Polarization in cadmium telluride nuclear radiation detectors," *IEEE Transactions on Nuclear Science*, vol. 23, no. 1, pp. 159–170, 1976.

[54] M. C. Veale, S. J. Bell, P. Seller, M. D. Wilson, and V. Kachkanov, "X-ray microbeam characterization of a small pixel spectroscopic CdTe detector," *Journal of Instrumentation*, vol. 7, pp. P07017–P07017, July 2012.

[55] P. Seller, S. Bell, R. Cernik, C. Christodoulou, C. Egan, J. Gaskin, S. Jacques, S. Pani, B. Ramsey, C. Reid, P. Sellin, J. Scuffham, R. Speller, M. Wilson, and M. Veale, "Pixellated cd(zn)te high-energy x-ray instrument," *Journal of instrumentation : an IOP and SISSA journal*, vol. 6, 12 2011.

[56] W. Shockley, "Currents to conductors induced by a moving point charge," *Journal of Applied Physics*, vol. 9, no. 10, pp. 635–636, 1938.

[57] S. Ramo, "Currents induced by electron motion," *Proceedings of the IRE*, vol. 27, no. 9, pp. 584–585, 1939.

[58] M. D. Wilson, P. Seller, M. C. Veale, and P. J. Sellin, "Investigation of the small pixel effect in cdznte detectors," in *2007 IEEE Nuclear Science Symposium Conference Record*, vol. 2, pp. 1255–1259, 2007.

[59] M. Veale, S. Bell, D. Duarte, A. Schneier, P. Seller, M. Wilson, and K. Iniewski, "Measurements of charge sharing in small pixel cdte detectors," *Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment*, vol. 767, 09 2014.

[60] A. Brambilla, C. Boudou, P. Ouvrier-Buffet, F. Mougel, G. Gonon, J. Rinkel, and L. Verger, "Spectrometric performances of CdTe and CdZnTe semiconductor detector arrays at high X-ray flux," pp. 1753–1757, IEEE, Oct. 2009.

[61] R. Ballabriga, M. Campbell, E. H. M. Heijne, X. Llopart, and L. Tlustos, "The Medipix3 Prototype, a Pixel Readout Chip Working in Single Photon Counting Mode With Improved Spectrometric Performance," *IEEE Transactions on Nuclear Science*, vol. 54, pp. 1824–1829, Oct. 2007.

[62] D. Pennicard, S. Smoljanin, F. Pithan, M. Sarajlic, A. Rothkirch, Y. Yu, H. Liermann, W. Morgenroth, B. Winkler, Z. Jenei, H. Stawitz, J. Becker, and H. Graafsma, "LAMBDA 2M GaAsA multi-megapixel hard X-ray detector for synchrotrons," *Journal of Instrumentation*, vol. 13, pp. C01026–C01026, Jan. 2018.

[63] J. Iwanczyk, E. Nygard, O. Meirav, J. Arenson, W. Barber, N. Hartsough, N. Malakhov, and J. Wessel, "Photon counting energy dispersive detector arrays for x-ray imaging," IEEE, 2007.

[64] M. C. Veale, P. Seller, M. Wilson, and E. Liotti, "HEXITEC: A High-Energy X-ray Spectroscopic Imaging Detector for Synchrotron Applications," *Synchrotron Radiation News*, vol. 31, pp. 28–32, Nov. 2018.

[65] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography," *Physics in Medicine and Biology*, vol. 21, pp. 733–744, Sept. 1976.

[66] Y. Lu, M. Kowarschik, X. Huang, Y. Xia, J.-H. Choi, S. Chen, S. Hu, Q. Ren, R. Fahrig, J. Hornegger, and A. Maier, "A learning-based material decomposition

pipeline for multi-energy x-ray imaging," *Medical Physics*, vol. 46, pp. 689–703, Feb. 2019.

[67] Jimenez, Edward Steven,, "Big-Data Multi-Energy Iterative Volumetric Reconstruction Methods for As-Built Validation & Verification Applications.," Tech. Rep. SAND2018-10707, 1475102, Sept. 2018.

[68] Z. Chen and L. Li, "Robust multimaterial decomposition of spectral CT using convolutional neural networks," *Optical Engineering*, vol. 58, p. 1, Jan. 2019.

[69] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2015.

[70] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv:1512.03385 [cs]*, Dec. 2015. arXiv: 1512.03385.

[71] M. Touch, D. P. Clark, W. Barber, and C. T. Badea, "A neural network-based method for spectral distortion correction in photon counting x-ray CT," *Physics in Medicine and Biology*, vol. 61, pp. 6132–6153, Aug. 2016.

[72] H. Yan, N. Bouet, J. Zhou, X. Huang, E. Nazaretski, W. Xu, A. P. Cocco, W. K. S. Chiu, K. S. Brinkman, and Y. S. Chu, "Multimodal hard x-ray imaging with resolution approaching 10 nm for studies in material science," *Nano Futures*, vol. 2, p. 011001, Mar. 2018.

[73] T. L. Sheppard, S. W. T. Price, F. Benzi, S. Baier, M. Klumpp, R. Dittmeyer, W. Schwieger, and J.-D. Grunwaldt, "In Situ Multimodal 3D Chemical Imaging of a Hierarchically Structured Core@Shell Catalyst," *Journal of the American Chemical Society*, vol. 139, pp. 7855–7863, June 2017.

[74] E. D. Martello, G. Tranell, O. Raaness, and L. Arnberg, "Combined XRD and XRF Technique for the Quantification of the Mass Balance in a Si Carbothermic Production Experiment," *ISIJ International*, vol. 51, no. 9, pp. 1492–1496, 2011.

[75] G. R. Pereira, H. S. Rocha, M. J. Anjos, P. Faria, C. A. Prez, and R. T. Lopes, "X-ray fluorescence and X-ray transmission microtomography imaging system," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 581, pp. 128–132, Oct. 2007.

[76] A. Rahman, G. Timms, M. S. Shahriar, C. Sennersten, A. Davie, C. A. Lindley, A. D. Hellicar, G. Smith, D. Biggins, and M. Coombe, "Association Between Imaging and XRF Sensing: A Machine Learning Approach to Discover Mineralogy in Abandoned Mine Voids," *IEEE Sensors Journal*, vol. 16, pp. 4555–4565, June 2016.

[77] D. Kasparek, B. McDonald, A. Gilbert, and R. Wittman, "Detector characterization for quantitative spectral radiography of uranium powder samples," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 954, p. 161413, Feb. 2020.

[78] A. Tomal, D. Cunha, M. Antoniassi, and M. Poletti, "Response functions of Si(Li), SDD and CdTe detectors for mammographic x-ray spectroscopy," *Applied Radiation and Isotopes*, vol. 70, pp. 1355–1359, July 2012.

[79] M. Van Gysel, P. Lemberge, and P. Van Espen, "Implementation of a spectrum fitting procedure using a robust peak model," *X-Ray Spectrometry*, vol. 32, pp. 434–441, Nov. 2003.

[80] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytracek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell'Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J. Gmez Cadenas, I. Gonzlez, G. Gracia Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampn, V. Lara, V. Lefebure, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. Mora de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O'Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. Di Salvo, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, M. Takahata, S. Tanaka, E. Tcherniaev, E. Safai Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wander, H. Weber, J. Wellisch, T. Wenaus, D. Williams, D. Wright, T. Yamada, H. Yoshida, and D. Zschiesche, "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, pp. 250–303, July 2003.

[81] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce Dubois, M. Asai, G. Barrand, R. Capra, S. Chauvie, R. Chytracek, G. Cirrone, G. Cooperman, G. Cosmo, G. Cuttone, G. Daquino, M. Donszelmann, M. Dressel, G. Folger, F. Foppiano, J. Generowicz, V. Grichine, S. Guatelli, P. Gumplinger, A. Heikkinen, I. Hrivnacova, A. Howard, S. Incerti, V. Ivanchenko, T. Johnson, F. Jones, T. Koi, R. Kokoulin, M. Kossov, H. Kurashige, V. Lara, S. Larsson, F. Lei, O. Link, F. Longo, M. Maire, A. Mantero, B. Mascialino, I. McLaren, P. Mendez Lorenzo, K. Minamimoto, K. Murakami, P. Nieminen, L. Pandola, S. Parlati, L. Peralta, J. Perl, A. Pfeiffer, M. Pia, A. Ribon, P. Rodrigues, G. Russo, S. Sadilov, G. Santin, T. Sasaki, D. Smith, N. Starkov, S. Tanaka, E. Tcherniaev, B. Tome, A. Trindade, P. Truscott, L. Urban, M. Verderi, A. Walkden, J. Wellisch, D. Williams, D. Wright, and H. Yoshida, "Geant4 developments and applications," *IEEE Transactions on Nuclear Science*, vol. 53, pp. 270–278, Feb. 2006.

[82] G. G. Poludniowski and P. M. Evans, "Calculation of x-ray spectra emerging from an x-ray tube. Part I. Electron penetration characteristics in x-ray targets: Calculation of x-ray spectra. Part I," *Medical Physics*, vol. 34, pp. 2164–2174, May 2007.

[83] "Charge sharing in silicon pixel detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 487, no. 1, pp. 113 – 122, 2002.

[84] M. C. Veale, J. Kalliopuska, H. Pohjonen, H. Andersson, S. Nenonen, P. Seller, and M. D. Wilson, "Characterization of M-π-n CdTe pixel detectors coupled to HEXITEC readout chip," *Journal of Instrumentation*, vol. 7, p. C01035, Jan. 2012.

[85] W. C. Gillis, A. J. Gilbert, K. Pazdernik, and A. Erickson, "A partial-volume correction for quantitative spectral x-ray radiography," *IEEE Transactions on Nuclear Science*, vol. 67, no. 11, pp. 2321–2328, 2020.

[86] M. Firsching, P. T. Talla, T. Michel, and G. Anton, "Material resolving x-ray imaging using spectrum reconstruction with medipix2," *Nuclear Instruments and Methods in Physics Research A*, vol. 591, pp. 19–23, 6 2008.

[87] X. Llopart, M. Campbell, R. Dinapoli, D. San Segundo, and E. Pernigotti, "Medipix2: a 64-k pixel readout chip with 55-um square elements working in single photon counting mode," *IEEE Transactions on Nuclear Science*, vol. 49, pp. 2279–2283, 12 2002.

[88] K. I. Ignatiev, G. R. Davis, J. C. Elliott, and S. R. Stock, "Microct (microtomography) quantification of microstructure related to macroscopic behaviour: Part 1 fatigue crack closure measured in situ in aa 2090 compact tension samples," *Materials Science and Technology*, vol. 22, pp. 1025–1037, Sept. 2006.

[89] D. J. Bull, I. Sinclair, and S. M. Spearing, "Partial volume correction for approximating crack opening displacements in cfrp material obtained from micro-focus x-ray ct scans," *Composites Science and Technology*, vol. 81, pp. 9–16, June 2013.

[90] J.-M. Kuhnigk, V. Dicken, L. Bornemann, A. Bakai, D. Wormanns, S. Krass, and H.-O. Peitgen, "Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic ct scans," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 417–434, 2006.

[91] A. J. Gilbert, B. S. McDonald, and M. R. Deinert, "Advanced algorithms for radiographic material discrimination and inspection system design," *Nuclear Instruments and Methods in Physics Research*, vol. 385, pp. 51–58, 2016.

[92] N. Collins, E. S. Jimenez, and K. R. Thompson, "Material identification with multichannel radiographs," *AIP Conference Proceedings*, vol. 1806, no. 1, p. 130004, 2017.

[93] P. B. Rose, A. S. Erickson, M. Mayer, J. Nattress, and I. Jovanovic, "Uncovering Special Nuclear Materials by Low-energy Nuclear Reaction Imaging," *Scientific Reports*, vol. 6, Apr. 2016.

[94] J. Harms, P. B. Rose, and A. Erickson, "Characterization of γ-ray cross talk in cherenkov-based detectors for active interrogation imaging applications," *IEEE Sensors Journal*, vol. 17, pp. 6707–6715, 10 2017.

[95] G. Beldjoudi, V. Rebuffel, V. Kaftandjian, and J. Rinkel, "Multidimensional data processing methods for material discrimination using an ideal x-ray spectrometric photon counting system," *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 3808–3815, 10 2010.

[96] J. Rinkel, G. Beldjoudi, V. Rebuffel, C. Boudou, P. Ouvrier-Buffet, G. Gonon, L. Verger, and A. Brambilla, "Experimental evaluation of material identification methods with cdte x-ray spectrometric detector," *IEEE Transactions on Nuclear Science*, vol. 58, pp. 2371–2377, 10 2011.

[97] E. S. Jimenez, K. R. Thompson, and L. J. Orr, "Utilization of virtualized environments for efficient x-ray attenuation approximation," *23rd ASNT Annual Research Symposium*, pp. 56–60, 3 2014.

[98] E. S. Jimenez, L. J. Orr, M. L. Morgan, and K. R. Thompson, "Exploring mediated reality to approximate x-ray attenuation coefficients from radiographs," vol. 9215, 8 2014.

[99] E. S. Jimenez, L. J. Orr, and K. R. Thompson, "Object composition identification via mediated-reality supplemented radiographs," *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 11 2014.

[100] E. S. Jimenez, N. M. Collins, E. A. Holswade, M. L. Devonshire, and K. R. Thompson, "Developing imaging capabilities of multi-channel detectors comparable to trditional x-ray detector technology for industrial and security applications," *SPIE Optical Engineering + Applications*, p. 99690A, 10 2016.

[101] B. De Man, *Iterative Reconstruction for Reduction of Metal Artifacts in Computed Tomography*. PhD, Katholieke Universiteit Leuven, Belgium, May 2001.

[102] W. Zbijewski and F. J. Beekman, "Characterization and suppression of edge and aliasing artefacts in iterative x-ray CT reconstruction," *Physics in Medicine and Biology*, vol. 49, pp. 145–157, Jan. 2004.

[103] C. R. Vogel, *Computational Methods for Inverse Problems*. Philadelphia: Society for Industrial and Applied Mathematics, 2002.

[104] R. T. Haftka and Z. Gurdal, *Elements of Structural Optimization*. New York City: Springer-Verlag, 3 ed., 1992.

[105] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York City: Springer, 2nd ed., 2006.

[106] P. Dvornyak, M. Koestlbauer, A. Lebrun, M. Murray, V. Nizhnik, C. Saidler, and T. Twomey, "Electrically cooled germanium system for measurements of uranium enrichments in uf6 cylinders (iaea-cn–184)," in *Symposium on International Safeguards*, International Atomic Energy Agency (IAEA), 2010.

[107] F. S. Voss and R. E. Greene, "A precise potentiometric titration method for the determination of uranium," technical report, 8 1953.

[108] N. Erdmanna, N. Alberta, P. Amadora, P. Arbora, H. Eberlea, K.Ltzenkirchena, H. Ottmara, H. Schorla, P. van Bellea, F. Lipcseib, P. Schwalbachb, S. Jungc, and R.Lafoliec, "Compucea 2nd generation performance evaluation (iaea-cn–184/316)," in *Symposium on International Safeguards*, International Atomic Energy Agency (IAEA), 2010.

[109] Y. Chen, C. Li, P. Ghamisi, X. Jia, and Y. Gu, "Deep Fusion of Remote Sensing Data for Accurate Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, pp. 1253–1257, Aug. 2017.

[110] H. Liu, T. Fang, T. Zhou, and L. Wang, "Towards Robust Human-Robot Collaborative Manufacturing: Multimodal Fusion," *IEEE Access*, vol. 6, pp. 74762–74771, 2018.

[111] V. Vielzeuf, A. Lechervy, S. Pateux, and F. Jurie, "Multilevel sensor fusion with deep learning," *IEEE Sensors Letters*, vol. 3, no. 1, pp. 1–4, 2019.

[112] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, pp. 423–443, Feb. 2019.

[113] J. Tobin, R. H. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 23–30, 2017.

[114] X. Zhou and S. Prasad, "Domain adaptation for robust classification of disparate hyperspectral images," *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 822–836, 2017.

[115] H. Emami, M. Dong, S. Nejad-Davarani, and C. Glide-Hurst, "Generating synthetic cts from magnetic resonance images using generative adversarial networks," *Medical Physics*, vol. 45, p. 36273636, 2018.

[116] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, 2018.

[117] G. Wilson and D. Cook, "A survey of unsupervised deep domain adaptation," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, pp. 1 – 46, 2018.

[118] A. A. Rusu, M. Vecerík, T. Rothörl, N. Heess, R. Pascanu, and R. Hadsell, "Sim-to-real robot learning from pixels with progressive nets," *ArXiv*, vol. abs/1610.04286, 2017.

[119] M. Yan, I. Frosio, S. Tyree, and J. Kautz, "Sim-to-real transfer of accurate grasping with eye-in-hand observations and continuous control," *ArXiv*, vol. abs/1712.03303, 2017.

[120] A. Elshamli, G. W. Taylor, A. Berg, and S. Areibi, "Domain adaptation using representation learning for the classification of remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 9, pp. 4198–4209, 2017.

[121] K. Bousmalis, A. Irpan, P. Wohlhart, Y. Bai, M. Kelcey, M. Kalakrishnan, L. Downs, J. Ibarz, P. Pastor, K. Konolige, S. Levine, and V. Vanhoucke, "Using simulation and domain adaptation to improve efficiency of deep robotic grasping," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4243–4250, 2018.

[122] J. Tobin, W. Zaremba, and P. Abbeel, "Domain randomization and generative models for robotic grasping," *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3482–3489, 2018.

[123] J. Cai, Z. Zhang, and H. Cheng, "Grasping novel objects by semi-supervised domain adaptation," in *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pp. 626–631, 2019.

[124] S. Iqbal, J. Tremblay, T. To, J. Cheng, E. Leitch, A. Campbell, K. Leung, D. McKay, and S. Birchfield, "Toward sim-to-real directional semantic grasping.," *arXiv*, 2020.

[125] F. Sadeghi and S. Levine, "$(\text{cad})^2$rl: Real single-image flight without a single real image," *ArXiv*, vol. abs/1611.04201, 2017.

[126] J. Yoo, Y. Hong, and S. Yoon, "Autonomous UAV navigation with domain adaptation," *CoRR*, vol. abs/1712.03742, 2017.

[127] J. Zhang, L. Tai, P. Yun, Y. Xiong, M. Liu, J. Boedecker, and W. Burgard, "Vr-goggles for robots: Real-to-sim domain adaptation for visual control," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1148–1155, 2019.

[128] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield, "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1082–10828, 2018.

[129] S. Thalhammer, K. Park, T. Patten, and M. Vincze, "Sydd : Synthetic depth data randomization for object detection using domain-relevant background," 2019.

[130] P. Schmitter, J. Steinrcken, C. Rmer, A. Ballvora, J. Lon, U. Rascher, and L. Plmer, "Unsupervised domain adaptation for early detection of drought stress in hyperspectral images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 131, pp. 65 – 76, 2017.

[131] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semisupervised transfer component analysis for domain adaptation in remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3550–3564, 2015.

[132] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Igum, "Generative adversarial networks for noise reduction in low-dose ct," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[133] Y. Xue, T. Xu, H. Zhang, L. Long, and X. Huang, "Segan: Adversarial network with multi-scale l1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, pp. 383–392, 2018.

[134] A. Madani, M. Moradi, A. Karargyris, and T. Syeda-Mahmood, "Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1038–1042, 2018.

[135] Z. Han, B. Wei, A. Mercado, S. Leung, and S. Li, "Spinegan: Semantic segmentation of multiple spinal structures," *Medical Image Analysis*, vol. 50, p. 2335, 2018.

[136] X. Yi and P. Babyn, "Sharpness-aware low-dose ct denoising using conditional generative adversarial network," *Journal of Digital Imaging*, vol. 31, pp. 655–669, 2018.

[137] A. Bentaieb and G. Hamarneh, "Adversarial stain transfer for histopathology image analysis," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 792–802, 2018.

[138] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical Image Analysis*, vol. 58, p. 101552, 2019.

[139] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, "Gans for medical image analysis," *Artificial Intelligence in Medicine*, p. 101938, 2020.

[140] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS14, (Cambridge, MA, USA), p. 26722680, MIT Press, 2014.

[141] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv*, Nov. 2014.

[142] J. Gauthier, "Conditional generative adversarial nets for convolutional face generation," tech. rep., 2014.

[143] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv*, May 2014.

[144] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," *ArXiv*, vol. abs/1512.09300, 2016.

[145] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, 2017.

[146] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), pp. 234–241, Springer International Publishing, 2015.

[147] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," pp. 2242–2251, IEEE, Oct. 2017.

[148] T. Zhou, P. Krähenbühl, M. Aubry, Q. Huang, and A. A. Efros, "Learning dense correspondence via 3d-guided cycle consistency," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 117–126, 2016.

[149] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," *ArXiv*, vol. abs/1703.00848, 2017.

[150] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2242–2251, 2017.

[151] F. Mahmood, R. Chen, and N. Durr, "Unsupervised reverse domain adaptation for synthetic medical images via adversarial training," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 2572–2581, 2018.

[152] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 95–104, 2017.

[153] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *arXiv:1611.07004 [cs]*, Nov. 2018.

[154] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[155] C. Wu, L. Herranz, X. Liu, Y. Wang, J. van de Weijer, and B. Raducanu, "Memory replay gans: learning to generate images from new categories without forgetting," *CoRR*, vol. abs/1809.02058, 2018.

[156] G. M. van de Ven, H. Siegelmann, and A. Tolias, "Brain-inspired replay for continual learning with artificial neural networks," *Nature Communications*, vol. 11, 2020.